The Segment in Phonetics and Phonology
### Contents

Notes on Contributors  

1 Introduction  
\textit{Eric Rainmy and Charles E. Cairns}

#### Part I  Is Segmentation Real?

2 The Segment in Articulatory Phonology  
\textit{Carol A. Fowler}

3 Beyond the Segment  
\textit{Markus A. Pöchtrager}

4 A Prosodic Theory of Vocalic Contrasts  
\textit{Chris Golston and Wolfgang Kehrein}

5 Segmentation and Pinky Extension in ASL Fingerspelling  
\textit{Jonathan Keane, Diane Brentari, and Jason Riggle}

6 Categorical Segments, Probabilistic Models  
\textit{Kathleen Currie Hall}

#### Part II  What Are the Roles of Segments in Phonology?

7 The Opponent Principle in RcvP: Binarity in a Unary System  
\textit{Harry van der Hulst}

8 Why the Palatal Glide Is Not a Consonant in Japanese: A Dependency-based Analysis  
\textit{Kuniya Nasukawa}
9 Determining Cross-Linguistic Phonological Similarity Between Segments: The Primacy of Abstract Aspects of Similarity
Charles B. Chang

10 Contrast and Vowel Features
San Duanmu

11 The Phonetics and Phonology of Segment Classification: A Case Study of /v/
Christina BJORNDahl

Part III Case Studies

12 The Perception of Vowel Quality and Quantity by Turkish Learners of German as a Foreign Language
Katharina Nimz

13 Compensatory Lengthening in Hungarian VnC Sequences: Phonetic or Phonological?
Mária Gósy and Robert M. Vago

14 Pári Consonant Mutation as Defective Root Node Affixation
Jochen Trommer

15 Templates as Affixation of Segment-sized Units: The Case of Southern Sierra Miwok
Eva Zimmermann

Subject and Author Index

Language Index
Notes on Contributors

**Christina Bjorndahl** is a doctoral candidate in Linguistics at Cornell University. Her dissertation research comprises a cross-linguistic phonetic and phonological study of the class of voiced, non-strident fricatives, with special attention on the segment transcribed as /v/.

**Diane Brentari** is Professor of Linguistics at the University of Chicago. Her current research interests include phonology, morphology, and sign languages, particularly as they relate to issues of typology, language emergence, and prosody. She has also developed a model of phonological structure of sign, called the Prosodic Model, and she has also worked on the architecture of the sign language lexicon.

**Charles E. Cairns** is Professor Emeritus of Linguistics at Queens College and the Graduate Center of the City University of New York. His research is in the general area of phonological theory. He is founder and, with Eric Raimy, co-organizer of the CUNY Phonology Forum, which has organized ten conferences since his retirement in 2003. Raimy and Cairns edited *Contemporary Views on Architecture and Representations in Phonology*, published by MIT Press in 2009. He and Professor Raimy also edited the *Handbook of the Syllable*, published by Brill in 2011; they also co-authored “Precedence relations in phonology,” which appeared in *The Blackwell Companion to Phonology* in 2011.

**Charles B. Chang** is a Lecturer in Linguistics at Rice University. Dr. Chang’s research addresses questions at the intersection of phonetics, phonology, and language contact. Recent and ongoing projects have examined the early stages of second-language phonological acquisition; interactions between the native and target languages in the mind of a second-language learner; language transfer; heritage language phonology; and contact-induced sound change.
San Duanmu is Professor of Linguistics at the University of Michigan. He received his Ph.D. in Linguistics from MIT in 1990 and has held teaching posts at Fudan University, Shanghai (1981–1986) and the University of Michigan, Ann Arbor (1991–present). His research focuses on general properties of language, especially those in phonology.

Carol A. Fowler is Professor Emerita at University of Connecticut and former President and Senior Research Scientist at Haskins Laboratories. Her research is on speech perception and speech production from an ecological perspective.

Chris Golston is Professor of Linguistics at California State University Fresno and a member of the Chukchansi Revitalization Project. His research is on phonology, morphology, meter, and the syntax-phonology interface.

Mária Gósy is a Professor at Eötvös Loránd University and Head of the Phonetics Department of the Linguistics Institute of the Hungarian Academy of Sciences. Her research areas are phonetics and psycholinguistics. Her current work focuses on the speech production process and on disfluencies of spontaneous speech.

Kathleen Currie Hall is an Assistant Professor in the Department of Linguistics at the University of British Columbia, where her research focuses on the perception, production, and modeling of phonological relationships, using techniques from a wide variety of areas, including experimental phonetics, psycholinguistics, sociolinguistics, semantics, and information theory.

Harry van der Hulst is Professor of Linguistics at the University of Connecticut. He specializes in phonology and has done research in feature systems and segmental structure, syllable structure, word accent systems, vowel harmony, sign language phonology, the phonology-phonetic interface, historical phonology, child phonology, language evolution, and cognitive science.

Jonathan Keane is a Ph.D. candidate in Linguistics at the University of Chicago. His current research interests are the phonetics and phonology of sign languages. Specifically the phonetics of fingerspelling, and how that informs handshape as well as articulatory models of language production. Additionally he is interested in instrumental data acquisition and articulatory phonology.

Wolfgang Kehrein works as an Assistant Professor of German and European Linguistics at the University of Groningen and as a Guest Researcher at the University of Amsterdam. His research is on phonology, the phonetics-phonology interface, and historical linguistics.

Kuniya Nasukawa undertook research for his Ph.D. in Linguistics at University College London and is Professor of Linguistics at Tohoku Gakuin University, Japan.
His research focuses on prosody-melody interaction in phonology, and he is editor of *The Bloomsbury Companion to Phonology* (Bloomsbury 2013) and author of *A Unified Approach to Nasality and Voicing* (Mouton 2005).

**Katharina Nimz** is a doctoral candidate in Linguistics at the University of Potsdam and the University of Newcastle-upon-Tyne. After a teaching position at the Humboldt University of Berlin, she is now working on her Ph.D. thesis on L2 sound acquisition. Her main interest lies in the relationship between L2 sound perception and production and the influence of orthography on cross-linguistic speech phenomena.

**Markus A. Pöchtrager** received his Ph.D. in Linguistics from the University of Vienna and is currently Assistant Professor in Linguistics at Boğaziçi University, Istanbul. His research interests include phonological (meta)theory, phonology-morphology interaction, linguistics as a cognitive science, Finno-Ugric and Scandinavian languages as well as Turkic languages.

**Eric Raimy** is a Professor in the Department of English and the Chair of the Department of Linguistics at the University of Wisconsin-Madison. His research is focused on modularity in grammar especially the role that representations play and the documentation and analysis of dialects of English spoken in Wisconsin. He collaborates and co-organizes the CUNY Phonology Forum with Charles Cairns.

**Jason Riggle** is an Associate Professor of Linguistics at the University of Chicago and Director of the Chicago Language Modeling Lab. He works on computational models of spoken and signed languages. Much of his research focuses on ways that models of learning and communication relate to the frequency of various linguistic patterns.

**Jochen Trommer** is a Heisenberg Fellow at the University of Leipzig and specializes in theoretical phonology and morphology, with a focus on micro- and macro-variation in lesser studied languages (Albanian, Uralic, Nilotic, and Kiranti). Currently his main interests are the learning of morphological segmentation, the role of tone in phonology and morphology, and the residue of non-concatenative morphology: polarity and subtraction.

**Robert M. Vago** is Professor of Linguistics at Queens College and the Graduate Center of the City University of New York and Chair of the Department of Linguistics and Communication Disorders at Queens College. His research concentrates on phonological theory, Hungarian linguistics, and first-language attrition.

**Eva Zimmermann** is a Ph.D. student at Leipzig University. Her research is in theoretical phonology, theoretical morphology, and its interfaces. Currently, she is
working mainly on non-concatenative morphology (especially lengthening, subtraction, non-concatenative allomorphy) and its analysis in phonological theory. An empirical focus in her work has been the non-concatenative morphology in several Native American Indian languages and the verbal agreement morphology in Kiranti and Broader Algic.
1 Introduction

Eric Raimy and Charles E. Cairns

1 Scope and background

The segment is an elusive entity. On the one hand it appears so intuitively obvious that it might appear strange to some that it warrants the attention of an entire volume, yet on the other hand it is not at all clear what it is, where it comes from, or whether or not the concept is entirely chimerical and thus a hindrance in our attempts to understand human language. This volume takes a deliberately eclectic approach to the study of the segment. The editors believe that theoretical, empirical, and methodological heterogeneity provides the most effective strategy to elucidate any entity that is difficult to pin down to the satisfaction of all researchers. This is the approach we took in Cairns and Raimy (2011a) in the investigation of the syllable, an entity with a similarly gauzy existence; the current volume parallels that study, mutatis mutandis.

To the extent that the segment has been explicitly discussed at all, there has never been consensus on what the segment is, or even if it represents a cognitively real entity of any kind. What level does it exist on, if it exists at all? Is it merely part of a more basic unit like the syllable? Most scholars agree that segments should be broken down into more basic units of some sort. But if subphonemic units are the elementary atoms of phonological description, are they reliably coterminal, producing an emergent segment? That is, is the speech stream neatly segmented syntagmatically, each segment defined as a bundle of distinctive features? These are questions that have not been resolved to the satisfaction of all scholars in the field.

One reason the segment continues to be such an elusive entity is that its boundaries lack any uniform and direct phonetic correlates. Hockett's Easter egg analogy is particularly appropriate here. Hockett (1955: 210) compares the relationship
between the segment source and the speech stream as that between a series of raw, but decorated Easter eggs and the resulting smeared mess obtained by running the eggs on a belt through a wringer. The original eggs end up being mashed together with only the presence of different colors reflecting the original source. Hockett says that the task for the inspector is only to identify the eggs and not to put the eggs back together. Making the analogy to phonology and phonetics, are there any tasks in phonology that require the identification of the eggs in the smashed output? Or, can phonology proceed without caring about how many and what colors of eggs there were, and where one ended and the other began? Or, do eggs even exist in the first place?

The remainder of this chapter is divided into three more sections. Section 2 provides a short overview of the history of the segment in the language sciences in order to identify the main issues at play. The third section explores some issues in contemporary views of the segment, and the final section provides a brief overview of the chapters in this book.

2 A short history of the segment

The assumption that the speech stream is divided into segments goes back to the earliest serious investigation into human language. From Panini’s rules of the sixth century BCE, which operated on discrete segments of speech, through most contemporary theories of phonology and phonetics, this assumption has prevailed consistently, sometimes implicitly and sometimes explicitly. This assumption has by no means been uncontested, as is reflected in this volume. There are multiple tensions that can be identified in the question as to whether or not the segment exists.

One tension was that between viewing the segment as defined by the features of which it is composed versus viewing the segment as an atomic whole. The former view appears to date back to Johann Conrad Amman (1692, see Ohala 2010) who proposed a hierarchically arrayed set of features. Erasmus Darwin (1803), grandfather of Charles Darwin, proposed that speech sounds are to be defined in terms of their distinguishing characteristics, thus subordinating the segment to the feature.

The most visible and productive work in nineteenth-century phonetics and phonology assumed atomic segments. For example, since its first publication by Passy in 1888, the International Phonetic Alphabet to this day conceptualizes segments as atomic elements; the “features” that do appear in the IPA are meant as attributes of segments and classes of segments, not as primitive constituent parts that can occur independently of one another.

The Neogrammarians made significant scientific progress using the atomic segment. The Neogrammarians discovered that sound change was based on segments; every occurrence of a segment, or class of segments, obeys the rules of change in that language. Verner’s law was a particularly compelling example of the power of the segment. Verner and the other linguists at the time were attempting to explore the
idea that all historical change is segment-based, that is, that sounds changed as classes of segments, not as a result of idiosyncratic changes in individual words. The existence of apparent exceptions to basic generalizations such as Grimm’s Law was an incentive to push the concept further. It was Verner’s (1877) insight, expressed in “Eine Ausnahme der ersten Lautverschiebung,” that a particular sub-class of segments escaped Grimm’s Law, namely the class of PIE voiceless stops when intervocalic and following an unstressed syllable in the same word. This was a major achievement of enlightenment science and it rested entirely on the assumption that laws operate at the level of phonology, which is autonomous from syntax and semantics, and on classes of segments viewed as unanalyzable wholes.

Significantly, the compositional view of the segment did influence important developments during the nineteenth and early twentieth centuries, despite the dominance of the atomic segment concept. For example, Alexander Melville Bell furthered the compositional idea in his 1867 publication of Visible Speech, a system of symbols each of which is composed of depictions of the activities of the relevant speech organs. Bloomfield (1914: 34) describes “nine typical tongue positions, three along the horizontal plane, front, mixed, and back, and three along the vertical, high, mid, and low,” strongly suggesting a componential analysis of segments. The concept assumed a central role in Prague School phonology, where its revival was inspired by Saussure’s central idea that language may be analyzed as a formal system based on the notion of distinctive oppositions. These oppositions became conceptualized as features which were the fundamental components of phonemes, eventually morphing into the concept put forth by Chomsky and Halle in the Sound Pattern of English (1968), where the phonetic string is neatly sliced into a sequence of segments, each of which consists of a bundle of binary features.

A second tension concerning the history of the segment was that between viewing rhythmic properties of speech, notably the syllable, as a fundamental unit and the segment is derived from it, or the other way around; this tension is well represented in the current volume. The former view can be traced back to Joshua Steele, who published Prosodia Rationalis in 1779; Steele took as his point of departure not segments, but rhythmic properties of speech such as tone, pitch, and cadence. This idea gained some traction during the nineteenth century, as witnessed in Sweet’s 1877 A Handbook of Phonetics. The idea that the fundamental units of speech are in its rhythmic properties, not segments, reached its fullest expression in the work of Herbert Raymond Stetson in the first half of the twentieth century. We will return to Stetson’s contributions below. But note for now that the concept of the segment as primary was dominant in the nineteenth century.

3 Contemporary issues concerning the segment

The tension that perhaps most shapes current debate about the segment concerns methodology and philosophy of science, and it is based on how one interprets empiricism in scientific inquiry. Put in a simplified form, the tension is based around
observable vs. covert structure in the data collected from the speech stream. A strong empiricism will require the existence of the segment to be based on strictly observable measurements in acoustic or articulatory data, while weaker versions of empiricism are more tolerant of qualitative observations of behavior or introspective judgments that point to covert structure. The stronger forms of empiricism seem to have flourished in recent decades as more and better technology for measuring acoustic and articulatory aspects of speech have become available. Ohala (2010) is representative of contemporary views on the question of how speech is represented in the brain:

The first, most candid, answer to this question is: we don’t know. Even such a fundamental issue as whether phoneme-sized segments are employed – or employed at all stages of encoding and decoding – has not been settled. There is an abundance of candidate answers given to the question as to how speech is represented mentally but until they have been properly evaluated they are just speculations. Within phonology the basic criterion applied in evaluating claims about mental representation of language is simplicity and such related notions as naturalness and elegance. But these are quite subjective and we learn from the history of science that the workings of the universe do not always coincide with pre-conceived human preferences.

Insofar as phonetic – or psychological studies (there is not always a clear distinction) – can shed light on the structure and processing of speech sounds in the mind of the speaker at some stages before the activation of muscle contractions and in the mind of the listener at some stages after the acoustic signal is transduced into an auditory signal, they may help us to discover other aspects of speech representation in the brain.

Ohala’s disdain for qualitative observations and hypothesis formulation and evaluation suggests that only strictly phonetic (or psychological?) evidence may be adduced in answering the question of whether or not segments exist. This is not an uncommon view and is likely one cause for the divergence between linguists who admit qualitative data for consideration and speech scientists who insist on only considering observable data. The question of whether to accept qualitative data represents the long-standing empiricist–rationalist debate. We give only a cursory sketch of this debate here; Chater et al (Forthcoming), provides a thorough genealogy of this dispute, especially as it pertains to linguistics.

A strong empiricist point of view rejects qualitative data based on speaker intuitions while a rationalist point of view is more likely to accept them (see Sprouse 2011 and Sprouse and Almeda 2012 for a comparison of quantitative vs. qualitative sources of data in syntax). Of course, phonologists of a rationalist bent require their theories to be empirically testable; they differ from the empiricists we are discussing only in that they are more accepting of qualitative data supporting covert structure.

The skepticism concerning accepting qualitative sources as data found an early expression in the work of Raymond Herbert Stetson. As a professor at Oberlin College he built his “Oscillograph Laboratory,” where he developed his
sylable-based theory of speech production. Influenced by Sweet (1877) and Sievers (1881), he based his theory entirely on the dynamics of speech production. An empiricist at heart, he spurned theoretical abstractions and focused his research exclusively on the physical record. Because the production of actual continuous discourse is characterized by hierarchically arranged rhythmic patterns, he naturally looked at prosodic structure for the basic, atomic units of speech. Stetson (1928) proposed the “chest pulse,” essentially the syllable, as the smallest unit in the prosodic hierarchy. Consequently he insisted, in true empiricist fashion, that human speech can only be understood if the scientist not only starts with an examination of the physical record, but at no point posits any covert structure; this methodology revealed the syllable to him. Stetson’s interpretation of empiricism and his conclusion that the syllable, not the segment, is the basic unit of speech go hand in hand. Segments are not evident in the physical speech stream, but he believed that syllables were. As he and subsequent scholars maintained, a strong behaviorism, one that requires an obsessive hugging of the physical ground, provides no evidence for the existence of the segment.

Stetson scorned American Structuralists like Bernard Bloch and Charles Hockett for their exclusive focus on segments chiding that, “… speech sounds are not a series of separate noises produced now this way, now that, and adjacent like the beads on a string; they are rather phases of the chest pulse, the syllable” (Stetson 1945: 6). It is important to note that both Bloch and Hockett soon abandoned the “beads on a string” view of segments. Bloch, in “A set of postulates for phonemic analysis” (1948), pointed out that features, as descriptors of individual speech organs, do not always turn on or off at the same time, rendering the “beads on a string” concept incoherent. Similarly, Hockett’s “Componential analysis of Sierra Popoluca” (1947) also proposes the removal “of the linearity assumption [i.e., the beads on a string perspective of segments] from among our working principles.” We will revisit this idea below, but note that both Bloch’s and Hockett’s rejection of the view of phonetic representations as a string of discrete segments was not in favor of Stetson’s idea of the syllable as primary, but rather in favor of a notion which can be viewed as a precursor of Autosegmental Phonology and of the theory of Articulatory Phonology.

Leonard Bloomfield agreed with Stetson about the existence and importance of the syllable, but he, like most structuralists, also admitted the existence of phonemes. In his famous book Language (1933), Bloomfield illustrates the concept of phoneme distinctiveness by means of a commutation test applied to the English word pin. The segments in pin can be proven distinct because of the appearance of each of them in other words. For example, pin starts with the same sound as pot, pin has the same vowel as bit and pin ends in the sound as ben. This is important because we consider arguments for the phoneme as de facto arguments for the segment. When Bloomfield (1933: 79) explains that “These distinctive features occur in lumps, or bundles, each one of which we call a phoneme,” we consider this as directly relevant to the existence of the segment. We can understand Bloomfield’s position on phonemes as distinct from Stetson’s rejection of them based on two factors. The first is that Bloomfield’s commutation test used to demonstrate the existence of phonemes tends more
toward a rationalist approach to data in language than Stetson’s strict empiricism. It is the more qualitative type of judgments, where pin and fin are different words and end in the same sound, that are part of Bloomfield’s commutation test that is the key difference here. American Structuralists accepted this as a legitimate form of data while Stetson did not.

The second factor distinguishing Bloomfield from Stetson is that the commutation test approach can be understood as developing a theory that relates distinct syllables (for he used only monosyllabic forms in illustrating his commutation test) to each other based on subsyllabic structure. If it is admitted that a syllable has some sort of internal structure, then the logical question is what constitutes this structure. Positing a segment is a natural conclusion if one accepts more qualitative data based on speakers’ intuitions such as Bloomfield’s commutation test. From the bulk of Bloomfield’s writings it appears that he accepts phonemes as unitary wholes and these distinctive features are characteristics that are cues to their perception. Bloomfield’s perspective resembled that of most other American Structuralists and was compatible with their “softer” behaviorism, one that allowed speakers’ judgments to play an empirical role. Their faith in segments was also based in part on the naïve assumption that segments would be revealed in the physical record, which was not available to them at the time.

Prague School phonology, which was largely contemporaneous with the developments described above in North America, was beyond the pale of even the most liberal empiricist. Although this school (as represented by Trubetzkoy 1939, see Anderson 1985) was functionalist, the concepts they developed have been incorporated into contemporary mentalist perspectives. Their view that segments are composed of features was heralded in the twentieth century by Saussure’s (1916) idea of defining linguistic units in terms of oppositions, widely seen now as cognitive entities although presented originally in functionalist terms. The Prague School adopted Saussure’s insight and showed that it is a potent tool in attaining insight into the sound structure of human language. According to the Praguean perspective segments are defined by the features that inhere in them; features are not simply the characteristics of phonemes that render them perceptible. This is a step further toward abstract representations. For example, it was the Prague School that invented the concept of the “archiphoneme,” which is an abstract segment defined by a set of features that lacks one or more features necessary to determine its phonetic realization, e.g., /N/, a placeless nasal which will acquire the place features necessary for production by means of context sensitive operations. Obviously, such a segment is not a measurable, observable entity in the real world; its only existence is as a participant in a formal (and presumably testable) theory of how phonological aspects of words are stored and processed.

Segments are entirely derived entities in classical Praguean phonology (in that they are defined in terms of their constituent features), but they remain as entities nevertheless. Each segment is a node in an interlocking system of distinctive oppositions defined by the features. The distinctive oppositions must be defined
at a system wide level, which means that the same phonetic segment may be composed of different distinctive features. This position has the potential to be overly abstract but in practice the phonetic content of a segment is used as only a partial guide to how distinctive oppositions are coded (see Dresher 2009 and Hall 2011 for contemporary discussions). Segments occupy a psychological space and are not to be found directly in the physical signal; they are abstract, atemporal entities.

The idea that features, not segments, are fundamental has remained a basic assumption of most modern theories of phonology. Chomsky and Halle (1968) describe lexical formatives (roughly morphemes) as represented by strings of consonants and vowels dubbed “segments,” which are in fact defined as bundles of distinctive features, distinguished from boundary segments by the feature [+ segment]. Of course, the segment is a clear concept within the formal framework of SPE, but the authors make no attempt to define it outside this system. More than merely rationalist, Chomsky and Halle assume a mentalist position in that the phonological grammar is a construct of the human brain. It follows from this that phonological entities are only constructs of the brain and thus difficult if not impossible to find in an acoustic stream. Of course, there must be some physical aspect of the acoustic stream that is perceived and gives rise to segments, but there is no necessary reason to assume that these physical aspects are immediately obvious or discrete; in fact, as is discussed at length in later chapters of this volume, there is evidence that listeners find some non-distinctive features to be perceptually salient.

The ultimate expression of segments as abstract markers of some kind occurred with the advent of autosegmental phonology (Goldsmith 1976, 1990; McCarthy 1989) and feature geometry (see McCarthy 1988 for a review). Autosegmental phonology was introduced into generative phonology in Goldsmith’s 1976 dissertation as “an attempt to supply a more adequate understanding of the phonetic” level of linguistic representation (1976: 16). It resembled contemporary versions of Articulatory Phonology (see Chapter 2 of this volume) in that it was “a theory of how the various components of the articulatory apparatus – the tongue, the lips, the larynx, the velum – are coordinated.” Goldsmith split the phonetic representation into several separate information channels, one for each articulator. Each channel contains information turning its respective articulator on or off; the entire set of channels represented what Goldsmith referred to as an “orchestral score.” Since articulators do not all turn off or on at the same moment, Goldsmith rejects the “Absolute Slicing Hypothesis,” the idea that phonetic representations can by cut up neatly into segments, each of which is a bundle of distinctive features. This allows for both many-to-one and one-to-many associations among features. Goldsmith (1976: 159) presents a diagram of an autosegmental phonetic representation as in (1), characterized as a segmentless phonetic score. Clearly, a rationalist/mentalist theory of phonology does not necessarily require obeisance to the Absolute Slicing Hypothesis, casting phonetic segments into doubt.
It is not obvious how to reconcile the graph in (1) with a discrete segment type object because the depicted object is not linear; each autosegment is a discrete object, as is each association line. Autosegmental, segmentless representations such as (1) were proposed to be at the phonetic level; Goldsmith agreed that segments must exist at the phonological level, although not all features inhere in segments at this level (i.e., some segments may “float,” to be attached by phonological rules or constraints). Thus, he said that there must be some sort of “deautosegmentalization” process to achieve SPE like feature matrices (1976: 159–165), given a phonetic representation like that in (1). A formal deautosegmentalization process has never been formalized and instead some sort of “skeleton” (Goldsmith 1990; Levin 1985; Broselow 1995) has been developed to provide a level of discreteness and ordering to autosegmental distinctive features. As explicated in Cairns and Raimy (2011b), skeletal slots play the main role of ordering and numerating “segments” in a phonological representation. The skeleton tells us how many segments there are and what the order is but not necessarily the content of each segment. Each segment is empty, in the sense that features do not inhere in them, but rather appear on autonomous planes and are associated with segments by means of either lexical association or rules of various kinds.

Phonologists who identify with rationalist traditions of generative phonology typically continue to adopt an unexamined notion of the segment. For example, none of the essays in the edited volume Lombardi (2001), Segmental Phonology in Optimality Theory, thought it necessary to define the segment or to question its existence. The essays in that book are concerned with the sorts of processes phonologists typically regard as involving segments, such as neutralization, vowel harmony, assimilation, and so on. Significant linguistic generalizations have been captured using the assumption that the segment exists.

It remains an empirical question within the rationalist tradition whether these generalizations might be better explained in a theory without the segment; this question is raised in later chapters of this volume. In contrast, contemporary phonologists more inclined to an empiricist point of view continue to suggest that it is wrong headed to entertain the segment in the first place. One school of thought holds that our tacit acceptance of the segment as a discrete entity is an artifact of literacy in an alphabetic language; Port and Leary (2005) and Silverman (2006) represent this position.

Both Port and Leary (2005) and Silverman (2006) argue that the rationalist view of the segment misrepresents the nature of phonology; they, along with Ohala, suggest that the generalizations captured by rationalist phonologists are illusory. Instead,
Introduction

Phonology is supposed to be concerned with a much finer grain of representation and include more detailed phonetic information than segment-based approaches allow. Port and Leary (2005) discuss both temporal information and the lack of complete neutralization in German devoicing and find no evidence for the segment. In our view, segments are atemporal in a formal theory of the mental representation of phonology, so Port and Leary (2005) are correct that the segment itself cannot be found in a temporal analysis of anything, no matter how meticulously detailed. In a similar vein, Silverman (2006) argues for an exemplar-based model of phonology, one which stores phonetically fine-grained utterances in exemplar clouds, one for each word. An exemplar model for phonology does not have segments though which raises the chicken and egg relationship about segments and alphabets. If segments do not exist at all, how did the idea of a segmental alphabet arise in order to lure contemporary phonologists to believe in the segment?

The idea that alphabetic writing systems have a causal relation with the existence of segments is well represented by Read et al. (1986) which investigates the ability to perform the metalinguistic task of adding or deleting a single consonant to a syllable among Mandarin speakers who varied on whether they were literate in an alphabetic writing system or not. The suggested conclusion of this study from the first line of the abstract is that, “Chinese adults literate only in Chinese characters could not add or delete individual consonants in spoken Chinese words” (Read et al. 1986: 31). Read et al. claim that segmentation is an emergent side effect of literacy in an alphabetic writing system or not. The suggested conclusion of this study from the first line of the abstract is that, “Chinese adults literate only in Chinese characters could not add or delete individual consonants in spoken Chinese words” (Read et al. 1986: 31). Read et al. claim that segmentation is an emergent side effect of literacy in an alphabetic writing system, presaging Silverman (2006) and Port and Leary (2005). This implies that there is no natural linguistic status for the category of segment. A closer look at these authors’ actual data, however, casts doubt on their conclusion. The nonalphabetic group did indeed perform much worse on the metalinguistic tasks; however, answers that added or deleted the non-targeted consonant were marked incorrect (Read et al. 1986: 38). If the erroneous addition or deletion of the non-target consonant were to be considered the manipulation of a segment along with the correct ones, a statistical reanalysis might yield different conclusions.

We wish to emphasize the truly elusive nature of the segment. We will see in a moment that morphophonemic and lexical phenomena seem to speak strongly for segment-like entities, but in fact a variety of psycholinguistic and phonetic studies give contradictory results. But no matter how one interprets the psycholinguistic findings about the existence of the segment, it would be a leap to conclude that the performance of subjects in psycholinguistic experiments can provide direct evidence for the existence of the phonological segment, because to do so would suggest that there is no role for the segment other than supporting the particular psycholinguistic tasks under investigation in the laboratory. Dunbar and Idsardi (2010) remind those who view phonology as a cognitive science that these metalinguistic tasks do not fall within the scope of the primary purpose (or even possibly a minority purpose) of the phonological segment. Instead, Dunbar and Idsardi (2010: 330) state that the goal of phonology should be to understand how, “... auditory and articulatory information is converted and consequently stored in and retrieved from long-term memory.” Under this conception of phonology, the positing of a segment seems
difficult to avoid, because to deny something akin to a segment would be to deny
the legitimacy of studying the structure of long-term memory.

The Dunbar and Idsardi perspective does not answer the question of what
constitutes a segment, but it does force the inquiry to be about what a segment's
characteristics are.

A primary question about the characteristics of a segment is its size. If the segment
is the unit of storage (Dunbar and Idsardi 2010: 330) for long-term memory then it
could be phoneme sized, syllable sized, foot sized, morpheme sized, and so on. This
really is the crux of the question. Relevant to this concern is Abler's (1989, 1997) argu-
ments that language follows the particulate principle, which is the idea that a complex
self-diversifying system is composed of small discrete units, particles.

Particulate systems are different from blending systems in what results from the
combination of parts of a complex system. A blending system operates like ink in
water, where there is a gradient in the overall color of the liquid based on how much
ink was added. Particulate systems on the other hand operate where the result of
combining two particles is something new and unrelated to the sum of them. Also,
the particles are recoverable in the combined form. Chemistry is an excellent
example of a particulate system. Sodium Chloride, NaCl is the result of the
combination of the metal sodium (Na) and the poisonous gas chlorine (Cl₂) but the
combination is common table salt. Abler (1989) argues convincingly that language
is a particulate system with multiple levels of organization; this is essentially the
same as Hockett's Duality of Patterning (Hockett 1960). Sentences are formed from
morpheme particles and morphemes are constructed from phoneme particles.

Abler's arguments have two important consequences for the segment. The first is
that the end of a particulate system is where a blending system is formed; some of
the debate about the existence of the segment can be understood by considering
that the phonology-phonetics interface is at this conversion point of language chang-
ing from a particulate system (i.e. segments) to a blending system (i.e. acoustics or
articulations). The second consequence is that particulate systems begin to behave
like blending systems when there are a large number of particles. This point suggests
that there is a premium on smaller numbers of fundamental particles as opposed
larger numbers. The number of phonemes required for describing any language is
typically considerably smaller than the number of syllables or other prosodic units
possible in that language, so the particulate principle favors the smaller number of
phonemes as a coding system as opposed to larger units like syllables, feet, and so on.

So we have good reason to suspect segment-sized entities are important in the
lexicon. In fact, it is hard to think of a coherent theory of phonological operations
that dispenses with segments at the lexical level. In the next few paragraphs we dem-
onstrate striking evidence that there must exist segments at some level of phonology.
Our argument is based on a morphophonological example from Temiar, a Mon-
Khmer language spoken in Western Malaysia by what Benjamin (2011: 5) describes
as “…an essentially pre-literate population.” The import of this observation is that
the process described below is a purely naturally occurring, ethologically valid
example of a segment-based, morphophonological process unrelated to a writing
system. Examples like this abound in the literature.
Consider the data taken from Benjamin (2011) in (2), focusing at first only on the material in slanting brackets, representing the lexical level, where word-formation rules such as those illustrated here operate. The base of the stem for “sit” is /gəl/, and the formation of the imperfective shows what Benjamin (2011) refers to as incopyfixation; part of the uninflected base form is repeated within the base as an inflectional marking. Monosyllabic forms, such as in (2a), reduplicate and prefix the first and last segments, gəl > ɡl-gəl. See Benjamin (2011: 2–3 fn. 4) for a fairly exhaustive list of analyses of this reduplication pattern.

As is typical of Mon-Khmer languages, Temiar also has sesquisyllabic forms such as in (2b). Benjamin (1976: 152–153) writes these types of forms phonetically with a vowel between the first two consonants, [səlog], but clearly argues that the written vowel is “… wholly determined by the absence of any other vowel … (p. 152)” thus positing a covert /slɔg/ representation. Yap (2006) argues that the lexical forms of sesquisyllabic stems contain an unsyllabified consonant followed by CVC syllable; the excrescent vowel is inserted by phonological rules or constraints not discussed here (see Benjamin 1976). Temiar forms the imperfective of sesquisyllabic stems by placing a copy of the last stem consonant immediately after the first stem consonant, səlog > s-g-əlog.

(2) Temiar incopyfixation

<table>
<thead>
<tr>
<th>Perfective</th>
<th>Imperfective</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. /gəl/</td>
<td>/glɡəl/</td>
<td>‘sit’</td>
</tr>
<tr>
<td>[ɡəl]</td>
<td>[ɡəɡəl]</td>
<td></td>
</tr>
<tr>
<td>b. /slɔɡ/</td>
<td>/sglɔɡ/</td>
<td>‘sleep, lie down, marry’</td>
</tr>
<tr>
<td>[səlog]</td>
<td>[sglɔɡ]</td>
<td></td>
</tr>
<tr>
<td>c. /ɡolap/</td>
<td>/bar-ɡolap/</td>
<td>‘to carry on, shoulder’</td>
</tr>
<tr>
<td>[ɡolap]</td>
<td>[bar-ɡolap]</td>
<td>[bə-ɡolap]</td>
</tr>
</tbody>
</table>

(2c) illustrates that stems that contain two or more lexical syllables form the imperfective by adding a prefix and do not undergo incopyfixation. Yap (2006: 83–94) specifically argues that a basic distinction between disyllabic (2c) and monosyllabic (2a and 2b) is necessary for fully understanding this particular allomorphy in Temiar. Forms which have two phonological syllables do not undergo incopyfixation while forms that only have one phonological syllable do.

The process of incopyfixation in Temiar provides strong evidence for the existence of a segment at some level of phonological representation. We place great credence in this evidence because it is naturally sourced and ethologically valid. Furthermore, there is strong evidence that it does not originate from the influence of an alphabetic writing system. Benjamin comments in various places on the status of literacy in the Temiar with the conclusion that, “… most of the data presented here therefore concern a language spoken by an essentially pre-literate population” (Benjamin 2011: 5); Benjamin collected his original data “…between 1964 and the late 1970’s” (Benjamin 2011: 3). The import of this observation is that incopyfixation can be considered a purely naturally occurring example of a segment-based morphophonological process unrelated to a writing system. Ethologically valid natural language phenomena such as the
incopyfixation data from Temiar clearly demonstrate that non-literate people have no problem acquiring and using segment-sized sounds in productive word-formation processes.

Experimental work concerning productive word-formation rules with preliterate children also provides strong arguments in favor of a segment not influenced by an alphabet. Jean Berko's famous Wug studies (Berko 1958) can be construed as evidence in favor of segments, because the formation of the regular past tense or plural in English involves adding a segment-sized linguistic entity, /d/ or /z/ respectively. Marcus et al. (1992) report that the most productive aspects of acquiring the English past tense are complete by the time children begin to read (the children studied achieve 90% + accuracy in marking regular past tense by the age of 3 years 6 months). This work strongly suggests that young children perform cognitive operations with segments without the influence of an alphabetic representation.

But phonological processes concerning word formation provide only one perspective on the segment. As we saw in discussing Read et al., the question remains open the extent to which fluency in an alphabetic writing system influences metalinguistic tasks of the sort investigated by Read et al. Nevertheless, we suggest that there is evidence in favor of the segment from contemporary research on lexical access. Samuel, in the Annual Review of Psychology (2011), states, “… a new consensus is forming. Taken together, recent research … indicates that the representations underlying spoken word recognition include (abstract) segmental codes along with prosodic and indexical information” (p. 61). This new consensus recognizes that multiple and distinct types of representations are necessary to fully account for lexical access, including segments.

Research also suggests that segment-size units play a role in early language acquisition, however illusive it may appear. Swingley and Aslin (2000: 161) say “we may say only that our results contradict any view holding both that (a) segmental representations are criterial for the differentiation of minimal pairs, and that (b) children of 18–23 months do not represent speech in terms of segments.” This study investigated the ability of children aged 13–23 months to notice small mispronunciations in words (e.g. vaby for baby). The experimental paradigm followed children’s eye movements while hearing a sentence that contained a target word. The results indicated that the children were able to recognize the mispronounced words but did so more slowly than comparable correctly pronounced words. This and other work with young, preliterate children are useful in teasing out the effects of the alphabetic writing system on metalinguistic tasks. This population also presents the opportunity to observe how segments may change over time or how lexical access develops sensitivity to different aspects of phonological representations. By accepting the existence of segments, we can focus on these types of investigations which will allow us to further refine our understanding of the role that segments play in different processes. If the segment is removed entirely from our nomenclature then these types of questions cannot even be formulated, not to mention studied.

As phonologists steeped in the rationalist/mentalist tradition, we conclude from the preceding that it appears difficult to dispense with the segment as some
sort of content-free, atemporal entity that is used in lexical storage and access, in word-formation rules, and is accessed by young children. However, we again emphasize two points: first, that there are rationalist phonological theories that make credible claims to legitimacy and dispense with the segment, as is demonstrated in this volume.

The second point we wish to emphasize is that in practically every domain of phonology and phonetics outside the lexicon the segment is not only difficult to pin down, there are psycholinguistic studies that produce directly contradictory data concerning the existence of the segment. Qu et al.'s (2012) ERP study of Mandarin Chinese is a good example of this. Both behavioral and ERP measures were made in a task that required subjects (native speakers of Mandarin Chinese) to name a picture specifying both an adjective and noun (e.g. green box). There were two conditions of stimuli that differed on whether the adjective and noun shared the same first segment (providing an opportunity for priming) or no phonological overlap between the two words. One result was that there was no behavioral evidence for the segment in that the two conditions did not produce any difference in naming latencies. Another result was that the ERP recordings differed between the two conditions based on time course and signal location, suggesting that there was a segmental effect present in the task. Thus, this single experiment produces conflicting evidence for the segment: the behavioral measures argue against the presence of segments in Mandarin, while the ERP measures argue for the presence of segments. One way of interpreting this situation is that Swingley and Aslin are correct when they suggest that segments can be present but at the same time not fully used in some fashion when accessing the lexicon. Presumably, one way that affects how robust behavioral effects based on segments appear is how alphabetic a writing system is.

The effect that segments are hard to identify using behavioral evidence in populations that are not highly literate in an alphabetic writing system may actually be too narrow an interpretation of this line of results. Huettig et al. (2011) investigated visual orientation behavior of both high literates and low literates in Hindi. Eye tracking was used to follow the gaze of subjects presented with four pictures in that consisted of one semantic competitor of the target word, one phonological competitor and two distractors that were not related either semantically or phonologically. The results of this study indicate that the two groups behaved differently in their gazing behavior. The high literate group's gaze initially fixated on the phonological competitor but then switched to the semantic competitor when it was clear that the pronounced target word was not the phonological competitor. The low literate group's gaze never fixated on the phonological competitor and slowly gravitated to the semantic competitor without reaching the same level of fixation as the high literates did. Eye tracking is a fairly automatic behavior so the difference we see in the two groups here strongly suggest that the general phonological processing is distinct in these two groups. Note that the low literate groups still must phonologically process the auditory stimuli because otherwise there should be no gaze focusing effect at all. There was no design aspect of the stimuli that would distinguish between segmental
and non-segmental representations, so we can only conclude that there is a general difference in phonological processing in high and low literates.

A general phonological processing difference that correlates with literacy is a possible novel interpretation of the effects discussed in Read et al.’s (1986) and Swingley and Aslin’s (2000) work. Both of these studies (and studies in a similar vein) demonstrate that a population that is not fully literate does not fully demonstrate the abilities for phonological manipulation that highly literate adults do. When this more general observation is combined with results from work like Qu et al. (2012) then we have a much more complicated picture of the role that a segment may play.

The fact is that we currently have conflicting evidence about the segment. We believe that the most productive way to organize the current set of knowledge about the segment is to look for emerging patterns in the conflicting evidence. The strongest evidence for the necessity of the existence of the segment comes from the more abstract levels of linguistic representation posited to explain word-formation processes in morphology such as the incopyfixation phenomena in Temiar and the child’s acquisition of the lexicon. These sources of data are primarily qualitative in nature so linguists of a more rationalist nature accept this type of data. The segment receives the least amount of evidence in favor of its existence and/or utility from very concrete levels of representation such as acoustic or articulatory phonetics or metalinguistic tasks. Linguists of a more empiricist nature favor this type of evidence and thus do not see strong arguments for the segment. Behavioral evidence for the segment correlates with level of literacy in an alphabetic writing system while neurolinguistic evidence is beginning to produce evidence in direct contrast to the behavioral data. Any single monolithic explanation for the segment (either affirming or denying its existence) will fail in providing an account of all of these different streams of data. Only a dynamic modular account of the segment will provide a framework which will be able to reconcile all of the data related to the segment that is accruing. This volume provides additional viewpoints on the segment with more specific arguments on where the segment is crucial in understanding some phenomenon and where it has little to no explanatory role. We believe that only by cataloging what phenomena require the segment and what phenomena do not will we be able to develop a comprehensive understanding of the segment.

4 The contents of this volume

The chapters of this volume are organized into three categories. The first extends the questioning begun in the previous section: what sort of existence claims can be made about the segment; the second inquires into the nature of segments in phonetics and phonology; finally, the third presents a number of detailed empirical studies of the nature and role of the segment.

Reality claims about the segment are examined in various ways in Part One, which is comprised of four chapters. Chapter 2 by Carol Fowler strongly supports the principle of duality of patterning and its attendant particulate principle, but questions
whether segments, as conventionally conceived, are the particles of interest. Arguing within the theory of Articulatory Phonology (AP), she argues that there is no level of phonology where strings are clearly segmentable. Whereas the classical Praguean segment consists of a bundle of (coterminous) features, AP posits gestures, which can overlap with each other along the time axis, thus providing no clearly demarcated point corresponding to a boundary. The meaningless particles that participate in the duality of patterning are gestures. AP does not envision segments along the physical time axis, because the articulators do not turn on and off all at the same time; nor are they atemporal, substance free entities. Fowler argues that AP accounts for many of the phenomena concerning phonologists thus calling into question the status of the segment. She also posits that there are other types of knowledge related to phonology and that this may be where the segment resides.

Marcus Pöchtrager, in Chapter 3, questions the traditional notion that the segment is the locus of contrastive differences between words. Applying Bloomfield’s commutation test to the American English words *bid* and *bit*, for example, provides no information about whether the voicing of the final consonant or the length of the vowel serves as the distinctive contrast. He examines the interaction of laryngeal contrast and vowel length in VC sequences in Estonian and English and presents a model where contrastive features are routinely independent of the traditional segment, thus down-playing the segment’s role in phonology. A core result of Pötrager’s reanalysis of English is that it resembles Estonian length distinctions, thus removing Estonian’s unusual position of having “overlong” vowels.

The reality of the segment is called into question from another tack by Chris Golston and Wolfgang Kehrein, in Chapter 4, where they argue that at least some distinctive oppositions inhere in prosodic nodes, not in segments. Their program is to reduce the status of the segment as a locus of distinctive oppositions, but their intent is to shift the burden to prosodic constituents in the lexicon. Golston and Kehrein specifically focus on vocalic features (i.e. rounding and palatalization) in syllable margins and demonstrate that the possible contrasts created by freely combining segment type representations vastly over generates the contrasts attested in human languages. Consequently, if vocalic features are directly associated to syllable positions, the remnant of a segment consists of only place and manner features.

So far we have only been considering segmentation in the acoustic domain. But spoken language is not the only mode representation available to humans. Keane, Brentari and Riggle in Chapter 5 present data demonstrating that fingerspelling, an important part of American Sign Language, shows segmentation issues entirely parallel to spoken language. It is not possible to impose neat segmentation boundaries into the stream of hand motion behavior produced in finger spelling, just as in speech. This is a particularly striking phenomenon because the input to this behavior obviously consists of a string of discrete representations. Although there exists debate about whether or not the representations underlying spoken language contain discrete elements, there can be no such debate regarding fingerspelling. Keane et al.’s work thus provides a clear example demonstrating the change from a discrete segment-like representation into a more gradient and fluid non-segmental structure.
Chapter 6 is the last of the first section of this volume and Kathleen Currie Hall argues that the debate about exclusively discrete vs. continuous representations is wrong headed. Instead, Hall shows how the phonemic vs. allophonic status of a phone can be determined by calculating different types of informational content over categories. This result clearly demonstrates that both types of representations, categorical and gradient, are necessary to fully understand phonology. Further extensions of the benefits of this approach are that Hall’s analysis provides insights into sound change, perception and variation.

The next five chapters comprise Part Two of the volume, and they inquire into the nature and roles of phonological segments. Harry van der Hulst, in Chapter 7 presents his model of phonological representations, Radical CV Phonology. Van der Hulst rejects the composition of segments from distinctive features and instead posits elements as the fundamental building blocks of contrast in phonology. The element approach is distinct from bundles of distinctive features in that each element has a direct complete pronounceable form where as a single distinctive feature does not. Van der Hulst explicates how elements map to phonetic forms and argues for a particular set of elements. The arguments about the number of elements answers element theory internal questions about how many primitives are required.

Chapter 8, by Kuniya Nasukawa, employs a strategy based on distributional facts to study the structure of glide segments. He shows that the palatal glide /j/ in Japanese is affiliated in terms of syllable structure with the following vowel as a light diphthong, not as conventionally thought as in the onset. This result has direct entailments for the internal structure of the glide segment; the author argues that the apparent sequence of a glide followed by a vowel is in fact a phonetic reflex of a single, complex segment defined by dependency relations among melodic primes. The light diphthong /ja/ consists of the same elements that make up the vowel /e/, but in different dependency relations; in the former, the element /I/, defining essentially palatality, is dependent on /A/, defining low vowel, whereas in the latter these dependency relations are reversed. As the author notes, this result will hold equally well in any theory of phonology that accepts segments.

Charles Chang, in Chapter 9, presents a thorough analysis of the various ways that speech sounds may bear similarity to one another in second language contexts. His study reinforces the disparity between phonetic and phonological types of similarity, showing that perceived cross-linguistic similarity of segments, as tested by judgments and other psycholinguistic tasks, is influenced more by phonological as opposed to phonetic factors. Thus, Chang argues that there is a strong role for an abstract segment type of representation in the understanding of second language phonology phenomena.

If it is assumed that segments are at least part of how words are stored in the human lexicon, then it is reasonable to ask if there are upper and/or lower bounds on the size of the inventory of segments. San Duanmu, in Chapter 10, shows that there are only 16 basic vowels in the universal inventory of contrastive segments. Basing himself empirically on an analysis of the database in UPSID and P-Base and using the Praguean notion of contrast as the main criterion for determining the
minimal number of phonemes, he argues that only four binary features are needed to distinguish vowels in the lexicon: [back], [high], [round], and [ATR].

Chapter 11 by Christina Bjorndahl, also shows the potency of cognitive segments in contrast to phonetic ones. Bjorndahl shows that the phonetic segment usually transcribed as [v] corresponds to different cognitive entities depending on the language in which it occurs, suggesting the existence of (at least) two distinct levels, the cognitive and the acoustic. A comparison between the phonetic and phonological behavior of [v] in three languages, Greek, Serbian, and Russian, demonstrates that [v] is the voiced counterpart of [f] in Greek, a sonorant in Serbian and of ambiguous status in Russian. Bjorndahl conducts an acoustic investigation of [v] in the three languages and shows that the [v] in Russian has a phonetic status similar to the [v] in Greek with Serbian [v] being distinct from both. If the phonetics directly indicate the phonology of [v] then Russian should behave like Greek but it does not. Bjorndahl’s cross-linguistic study suggests instead that phonological identity is determined at least in part by cognitive properties, such as contrastive and distributional patterns in the language.

Part Three of this volume contains four chapters, each a closely detailed empirical study of a rather narrow range of phenomena which throw light on the nature of segments. Katherina Nimz, in Chapter 12, approaches the segment concept from the field of second language phonetics and phonology. Nimz reports on experiments with native speakers of Turkish who are learning German. She concentrates on vowels, where the two languages’ vowel systems differ in terms of the use of both contrastive length (present in German, absent in Turkish except in borrowings and derived words) and quality (the languages have very different vowel inventories). It follows from her study that cross-linguistic vowel perception is guided by salient phonetic features even if that particular feature is not contrastive in the native language: contrastive length is present in German but not Turkish, yet Turkish listeners rely as much on duration as a cue in vowel perception as much as German speakers do. The conclusion is that a feature need not be contrastive to be salient, even among adults.

Mária Gósy and Robert M. Vago’s Chapter 13 offers a striking case study of the interface between the gradient phonetic world and the discrete phonological one in compensatory lengthening in Hungarian. Most analysts view compensatory lengthening as operations performed at some phonological level that contains ordered sequences of discrete phonemes. However, Gósy and Vago demonstrate that there are robust gradient aspects of compensatory lengthening in Hungarian causal speech. This provides an important example where non-segmental analysis is required for one to fully understand a segmental process. Both approaches are required.

To complete the volume, the final two chapters contain detailed phonological analyses of morphological operations in two languages that reveal interesting properties of and possible constraints on such lexical information. Jochen Trommer, in Chapter 14, and Eva Zimmermann in Chapter 15, explore proposals for underspecified segments and unaffiliated features. Trommer presents a detailed analysis of the morphophonology of Päri which requires a representation
which is an unspecified segment. This is possibly the most abstract form of a segment because it lacks any sort of distinctive feature content but Trommer illustrates the necessity of this aspect of the analysis of Päri.

Zimmermann’s analysis of Southern Sierra Miwok in Chapter 15 complements Trommer’s points in that an underspecified segment type representation is necessary. Furthermore, Zimmermann shows that one does not need to trade off representations for one another in that Southern Sierra Miwok requires both segmental and prosodic representations to fully explain its templatic aspects. This allows us to see what segmental structure contributes to the templatic structure of Southern Sierra Miwok and what it does not.

The overall constellation of the chapters in this volume provides an excellent overview of what the segment does and does not do for our understanding of phonology. No one chapter presents the whole picture but we believe that when all of them are put together the controversial and unsettled nature of the segment becomes apparent. Each chapter holds different assumptions and considers a different aspect of phonology so it is not surprising that different aspects (or non-aspects) of the segment emerge in each. We hope that a side effect of required point of view on the segment is more ecumenical cross disciplinary work in phonology.

Acknowledgments

The authors would like to thank all of the participants in the CUNY Phonology Forum Conference on the Segment held in January of 2012 and especially the contributors to this volume. The presentations and especially the chapters presented here have greatly deepened the authors’ understanding of the nature of the segment in phonology. We are particularly grateful for John Goldsmith’s contributions. All errors of fact and/or interpretation in this introduction are ours alone.

Notes

1 Although Stetson’s physiological theory of the syllable based on chest pulses was refuted by Ladefoged (1967) and others, Krakow (1999) presents arguments that the syllable is based on a legitimate physiological unit.

2 It is interesting to note that contemporary use of the phrase “beads on a string” is now a positive way of speaking about segments in phonology, which was not the original intention of Stetson.

3 Many thanks to John Goldsmith, p.c., for pointing out these references to us.

4 We are not interested in a subtle debate about the differences between a segment vs. a phoneme so we will treat the two as synonymous for the purposes of our discussion.

5 At least they were when the data were collected; Benjamins reports “in the five decades since I began researching their society and language, the people have undergone a huge shift – from isolated, non-literate tribal society to literate, physically mobile peasants and proletarians” (Benjamin 2011: 21).
References


Part I

Is Segmentation Real?
1 Introduction

This chapter is about the status of the segment in articulatory phonology. However, because articulatory phonology is not widely identified as a useful approach for addressing phonological issues, I begin by suggesting why a phonology should be articulatory. Likewise, because articulatory phonology may not be as familiar to readers as other approaches, I devote a little space to describing and defending the primitive phonological forms of that theoretical approach, emphasizing that, in articulatory phonology, phonological language forms are adapted to their public use. Articulatory phonology is the only phonology in which implementation of phonological forms in coarticulated speech is understood to be nondestructive of the forms' essential nature.

2 Why should a phonology be articulatory?

For the most part, the fields of linguistics and psycholinguistics have embraced the idea, made forcefully by Noam Chomsky, that the only scientifically interesting language domain is “I-language,” or internal language. For example: “The concept of E [external]-language, however construed, appears to have no significance” (Chomsky 1986: 31, italics added). Or, more recently and somewhat less radically: "Externalization by the SM [sensorimotor] system appears to be a secondary property of language" (Chomsky 2011: 275). That point of view certainly deflects attention from the observation that language serves a major role in interpersonal communication and from any idea that it may be adapted to that role. Most relevantly here, it does not naturally
reinforce an idea of articulatory phonologists that the phonological forms of languages (consonants and vowels and the forms they compose) should be articulatory (e.g., Browman and Goldstein 1986, 1992; Goldstein and Fowler 2003).

However, it is clear that language is an evolutionary adaptation of our species, and that its public implementation as speech is part of the evolved system. Although research in the last 30 years or so (e.g., Klima and Bellugi 1979; Reagan 2011) has established that the signed languages used largely by communities of deaf speakers have all of the expressive power of spoken languages, nonetheless, 100% of human cultures have spoken languages. It is not a coin toss whether the language of any hitherto unencountered human society will turn out to be spoken or signed.

Language forms are the means available within languages for making linguistic messages public and therefore available to be shared among language community members. If language is adapted for public use, and if language forms are the means within language for sharing linguistic messages, we should expect language forms to be adapted to implementation by the vocal tract and to perception from acoustic signals. And they clearly are in some respects.

In the languages of the world, small vowel inventories most frequently consist of vowels that maximize their distance apart in acoustic space (e.g., Lindblom and Maddieson 1988; Diehl and Kluender 1989). This provides evidence that perceptual distinctiveness shapes vowel inventories. Consonant inventories less obviously favor perceptual distinctiveness, but Lindblom and Maddieson (1988) show that language communities favor consonants that are easy to say. They partitioned consonants into three categories along a complexity cline: basic, elaborated, and complex, and they showed that languages with the smallest inventories have mostly the articulatorily simplest (basic) consonants. Only languages with the largest inventories have complex segments. Although vowel inventories most saliently favor perceptual distinctiveness, and consonants favor articulatory ease, Lindblom and Maddieson (1988) suggest that both types of segment inventory reflect both constraints. The important point here is that public language use, in which listeners need to apprehend what talkers say, and in which talkers need to speak fluently, shapes the sound inventories of language. The inventories are, as it were, designed for public use.

The phonologies of languages encompass more than segment inventories. Across the lexical inventories of languages, word forms exhibit regular properties that are captured by rules in the generative phonology of Chomsky and Halle (1968) and by constraints in Optimality Theory (Prince and Smolensky 2004). Some of these regular phonological processes appear to be conventionalizations of more general phonetic dispositions. For example, some languages have final devoicing “rules” in which voiced consonants surface as unvoiced when they are word final. This is likely to be a conventionalization of the general difficulty of sustaining voicing in final position (Westbury and Keating 1986). Likewise, some languages, for example, Hungarian and Turkish, have vowel harmony “rules” in which vowels in words share featural properties such as frontness-backness (e.g., Turkish, e.g., Polgárdi 1999; Hungarian, e.g., Benus and Gafos 2007) or rounding (e.g., Turkish). All or most languages exhibit vowel-to-vowel coarticulation, a likely phonetic source of
phonological vowel harmony. Consonant harmony is much less frequent across languages (e.g., Gafos 1999), and consonant-consonant coarticulation across a vowel is, for the most part, impossible.

Given these indications that the phonologies of languages are shaped by the public use of language, some conventional ideas about phonological language forms ought to be questioned. Among those questionable ideas is one that phonological language forms have their primary home in the mind of language users. In this idea, they are categories in the mind that are abstracted from their modes of implementation; they are cognitive rather than physical entities (e.g., Pierrehumbert 1990). They are supposed to be abstract in being discrete one from the other, static (in consisting of collections of featural attributes), and context-free. These properties apparently contrast with vocal tract actions during speech. Because speakers coarticulate when they speak, vocal tract actions that implement sequences of consonants and vowels are not obviously segmentable into discrete components. Nor are they described realistically as a sequence of static configurations with each configuration representing one phonological segment. Moreover, to the extent that movements of the articulators can be associated with individual segments, the movements are highly context-sensitive. In short, coarticulation is judged to be destructive (e.g., Hockett 1955) or distorting (Ohala 1981) of the fundamental character of phonological segments as language users are believed to know them.

Articulatory phonologists reject the idea that there is a fundamental incompatibility between properties of primitive phonological entities and vocal tract actions. As such, they hold that humans are adapted not just to learning and knowing language, but also to using it by speaking and listening. In articulatory phonology, language forms are adapted to their public use in every respect.

3 Articulatory phonology

In articulatory phonology, language forms have their primary home in the vocal tract, not in the mind. This is not to say that proponents suggest that language users are empty-headed, phonologically speaking. Rather, it is to say that language forms are public actions of the vocal tract. The mind may store whatever language users know about those actions. (A possibly helpful analogy: there are elephants in the world. Humans with direct or indirect experience of elephants may have knowledge of elephants in memory, but they do not have elephants in memory.)

Because language forms are fundamentally articulatory, they are adapted to public use. They are discrete one from the other in being separate actions of the vocal tract. As produced, they are not discrete in time; rather, they overlap temporally, but this does not eliminate their separateness; they are separate actions implemented by separate synergies of the vocal tract in overlapping time frames. They are fundamentally dynamic, not static. Finally, despite coarticulation (temporal overlap), their essential properties are context-free. Coarticulation is not distorting of or destructive of essential properties of phonological language forms.
In articulatory phonology, the smallest language forms are “phonetic gestures.” They are coarse-grained actions (that is, not generally movements of individual articulators) that create and release constrictions. Accordingly, their defining properties are the location in the vocal tract in which the constriction is made (“constriction location”) and the extent or degree of the constriction, from wholly closed (e.g., lip closing for /b/) to wide (e.g., for /a/).

Phonetic gestures are modeled as dynamical systems (e.g., Saltzman and Munhall 1989), which provide useful models of coordinated actions quite generally (e.g., Kelso, 1995; Turvey 1990; Warren 2006). The systems represent synergies of the vocal tract consisting, prototypically, of two or more articulators organized transiently to achieve a constriction in a particular location to a particular degree. As synergies, they exhibit “equifinality” – that is, a tendency to achieve the invariant constriction goals in variable ways. Synergies are revealed by experiments in which an articulator is perturbed during speech, and compensatory actions are observed. For example, Kelso, Tuller, Vatikiotis-Bateson, and Fowler (1984) perturbed the jaw by pulling it down on a small proportion of trials while the jaw was raising for the second /b/ in /baeb/ or for the /z/ in /baez/ produced in the carrier phrase: It’s a ______ again. Within 20–30 milliseconds of the onset of the perturbation, Kelso et al. observed extra activation of a muscle of the lip, and extra downward motion of the upper lip just when the perturbation was during /b/. The extra motion of the upper lip compensated for the displaced jaw position so that the lips achieved closure for /b/. This kind of compensatory action was not appropriate for /z/, and did not occur in production of that segment. Rather, the tongue (as indexed by extra activity in the genioglossus muscle, and by the audible frication) compensated for the lowered jaw so that an acceptable /z/ was produced.

These findings and many others (e.g., Abbs and Gracco 1984; Munhall, Löfqvist, and Kelso 1994; Shaiman 1989) verify that synergies or dynamical systems underlie speech production just as they underlie coordinated actions quite generally. However, their raison d’être cannot be to protect talkers against highly unlikely external sources of perturbation such as the “jaw puller” of Kelso et al. An important role outside the laboratory is to permit equifinal achievement of required constriction locations and degrees despite perturbations induced by coarticulating segments. In a syllable /bi/, for example, the jaw can adopt a high position that fosters lip closure for /b/ both because such a position fosters lip closure, but also because it fosters production of the high vowel /i/. However, in /ba/, the coarticulating open vowel /a/ will pull the jaw down during closing for the /b/, and this necessitates compensatory action by the lips.

Figure 2.1 (from Lindblom and Sussman, 2012) shows equifinal achievement of alveolar closure by the tongue tip during /d/ despite coarticulatory overlap by the following vowel. The figure shows X-ray tracings of sagittal views (right facing) of the vocal tract at two points in time during production of /da/and /di/. In the figure, the plain tongue contour reflects the position of the tongue during /d/ closure. The dotted contour reflects the position of the tongue during the
following vowel midpoint. Two observations are important for present purposes. One is that, across the syllables, there is invariant achievement of the constriction location and degree for /d/. The second is that this is equifinal achievement of /d/’s defining properties. The rest of the tongue contour in both syllables is quite different during /d/ closure, and it clearly reflects coarticulatory overlap of the consonantal and vocalic tongue gestures. That is, during /d/ closure, the contour of the tongue behind the tip clearly is on the way to its contour during the vowel midpoint.

4 Gestural scores

In articulatory phonology, gestural scores display the gestures for word forms of a language. Figure 2.2 shows a schematic gestural score for the word palm. Each tier of the score shows the articulator systems that implement gestures in the word. The component gestures of palm are shown by shaded blocks indicating the temporal intervals during the word in which the gestures exert an influence.
over vocal tract actions. Descriptions in the shaded regions specify the constriction locations and degrees of the gesture. The temporal overlap of the gestures across the different tiers shows the phasings among the gestures for *palm*. The parabolic curve in the glottal gesture is meant to show that the phasing of that gesture with the bilabial constriction gesture for /p/ is such that peak glottal opening coincides with bilabial constriction release so that the /p/ is aspirated as it is in that context in English.

The gestural score for a word might be considered to replace a lexical entry in a generative phonology in which segments are represented by columns of featural attributes. Notable differences with those representations are that gestural scores represent relative time, not just serial order along the horizontal dimension and that gestures overlap in time. Another notable difference is that segments are not always apparent in a gestural score. Two of the segments in *palm*, the /p/ and /m/, are two-gesture segments. /p/ includes a bilabial closure gesture and a glottal opening gesture. But nothing in the score ties those two gestures together any more than either is tied, say, to the vocalic gesture (the gesture having a pharyngeal constriction location of the tongue body and a wide constriction degree) with which both overlap in time. Likewise, nothing ties the velar opening and lip closing gestures at the end of the word despite the fact that, together, they are supposed to constitute the final consonant of the word.

The absence of a transparent relation between gestures and segments can also be seen in Figure 2.3, which presents schematic gestural scores for the words *sod* and *pod* and the nonword *spod*. Both /s/ and /p/ are associated with a glottal opening gesture as well as an oral constriction gesture as shown in the scores in the upper half of Figure 2.3. However, when they participate in a consonant cluster, only one glottal gesture remains. There is no obvious way to partition the cluster into distinct unvoiced consonantal segments.
5 Does it matter that segments do not emerge transparently in articulatory phonology?

A criticism of articulatory phonology might be that it is not really a phonology. Rather, it provides a useful phonetic description of speech. In that case, perhaps it does not matter that segments do not always emerge as obvious units of those descriptions.

However, Browman and Goldstein (e.g., 1986, 1992) did intend to provide a novel phonology, not just a novel phonetics. Of course different theoretical phonologies have differed in the issues on which they focus, giving articulatory phonologists a variety of issues to address. In the phonologies within descriptive linguistics (e.g., Gleason 1961) that antedated the generative phonology of Chomsky and Halle (1968), a major aim was to develop procedures for classifying phonetic segments into the same or different phonemic categories based on whether the featural distinctions between members of segment pairs were noncontrastive or contrastive. Chomsky and Halle jettisoned that effort, and focused on developing sets of rules to capture systematic phonological processes that span the lexicons of languages. More recent optimality theoretical approaches express those regularities as constraints.

Browman and Goldstein (e.g., 1992) showed that, in articulatory phonology, contrast can be captured gesturally. For example, /pat/ and /hat/ differ in the

Figure 2.3 Gestural scores for pod, sod, and spod.
presence/absence of a bilabial gesture, /bɪd/ and /dɪd/ differ in the constriction location of their oral constriction gesture, and /tɪk/ and /sɪk/ differ in the constriction degree of the alveolar gesture of the tongue tip for the initial consonant.

Although they have not explored whether articulatory phonology offers insight into systematic phonological processes, Gafos (e.g., 1999, 2002; Gafos and Benus 2006) has. For example, he has shown for Moroccan Arabic (Gafos 2002) that some processes or constraints in its phonology must refer to temporal relations among gestures. That is not a possibility in any phonology except articulatory phonology. Second, Gafos (1999) remarks on the striking difference previously mentioned in the popularity of vowel harmony and consonant harmony as phonological processes in the languages of the world. As noted earlier, vowel harmony is likely to emerge from the articulatory process of vowel-vowel coarticulation, a process that is language general because aspects of vowel production can typically be maintained during intervening consonants without preventing achievement of consonantal constriction goals. In contrast, for the most part, consonant production cannot persist during intervening vowel production without blocking the acoustic impact of vowel production. Not only is consonant harmony relatively rare, but when it occurs, typically (but perhaps not always, Hansson 2007), it involves consonantal parameters of tongue tip orientation or constriction area that can be maintained during a vowel. Third, Gafos and Benus (2006) have noted that implementation of some phonological processes (in their example, final devoicing in German) can be gradient rather than categorical. This is unexpected both from a rule conception of the processes (because rules either apply or do not) or from Optimality Theory’s constraint ranking approach (because outranked constraints are wholly invisible). Gafos and Benus suggest that the dynamical systems modeling of articulatory phonology advocated by articulatory phonologists provides a natural way for ranked constraints (ranked markedness and faithfulness constraints for final devoicing in German) to exhibit gradience.

In short, the gestures of articulatory phonology offer a realistic kind of language form if languages are understood to be adapted for public use. Addition or deletion of gestures or change in a gestural parameter value can capture linguistic contrast. Moreover, the intrinsically temporal character (Gafos 2002) of the gestures, that the gestures are themselves vocal tract actions (Gafos 1999), and that they are naturally modeled as dynamical systems (Gafos and Benus 2006) suits them also for characterizing fundamental systematic processes in the languages of the world. Therefore, it does matter in addressing the question of the existence of segments in language, whether segments emerge as phonological units in articulatory phonology.

6 Arguments for segments and articulatory phonology

For Browman and Goldstein (1990), the reality of conventional phonological segments is an empirical issue. Moreover, they are not optimistic that the issue will be resolved in favor of segments. “[W]e would argue that the basis for [segments]
seems to be their utility as a practical tool rather than their correspondence to important informational units of the phonological system” (p. 418).

For my part, however, I wonder how segments can have practical utility if they do not fairly closely reflect reality. Before looking at the most relevant empirical data from articulatory phonologists, I consider some well-known considerations that have inclined investigators and theorists to a view that segments are real. Doing without segments would require revising our understanding of these issues and findings.

7 Duality patterning and the particulate principle

Hockett (1960) proposed “duality of patterning” as one of the design features of language. The label refers to the fact that, in language, meaningful language forms, morphemes and words, are composed from a small stock of meaningless segments that Hockett identifies as “sounds” (p. 6). In his example, the words *tack*, *cat*, and *act* are composed of the same three meaningless sounds in different permutations, but they have wholly distinct meanings.

Relatedly, Abler (1989; cf. Studdert-Kennedy 1998) notes that language, like just a small number of other “self-diversifying” systems (genetic inheritance and chemical compounding) reflects effects of a “particulate principle.” These systems can make infinite use of finite means (von Humboldt 1836/1972), because their atomic particles combine without blending. Unlike the ingredients of a cake, for example, the atoms composing *tack*, *cat*, and *act* combine without losing their identities. In this way, different combinations of the same limited pool of atoms compose different molecular entities.

By most accounts, phonetic segments are the particles of language. If they are not, then what are the particles? In articulatory phonology, they must be the primitive gestures of that theory. However, if gestures are the particles underlying duality of patterning and the particulate principle, then we must turn to phonotactic constraints to understand why the particles are restricted in how they can combine so that they appear to compose segments. For example, in English, a velum lowering gesture can combine with a labial or alveolar or velar closing gesture, but only if there is no associated devoicing gesture.

8 Spontaneous errors of speech production

Spontaneous errors of speech production have served as an important source of evidence that language units identified as such in linguistic analysis are units as well for language users. When a speaker is heard to say *heat seater* intending to say *seat heater* or to say *beef needle soup* (Dell 1986) intending *beef noodle soup*, the errors appear to show that phonological segments are discrete combinable units of language planning. Are there other interpretations?
Browman and Goldstein (1990) argued that most segment errors can also be identified as feature errors and, therefore, gesture errors. For example, heat and seat differ by one segment, but also by one feature or gesture (place of articulation or presence/absence of a tongue tip gesture). Errors classified as segment errors characteristically are featurally similar and so might instead be errors involving one or two gestures.

Apparently opposed to their interpretation is one observation and one research finding. The observation is that, although many, perhaps even most, sublexical errors are ambiguously feature or segment errors (or even, syllable-constituent (e.g., onset, nucleus, coda) errors), there are few instances of errors that are unambiguously feature errors reported in transcription-based error collections. A well-known example is Fromkin's (1973) glear plue sky for clear blue sky. This is almost certainly an exchange error. (That is, it is almost certainly not two noncontextual substitutions: /g/ for /k/ and /p/ for /b/.) If it is an exchange error, the exchanging elements are either voicing features or the presence/absence of a devoicing gesture. Either way, the elements involved are subsegmental. However, again, such errors are infrequently reported in transcription data.

The finding is by Shattuck-Hufnagel and Klatt (1979). They explicitly addressed the question whether sublexical errors were most likely to be whole segment or else feature errors. They looked at a set of 70 exchange errors in their corpus in which the exchanging consonants differed by at least two phonetic features (place, manner, voicing). Then they could ask whether what exchanged was a whole segment (with all features moving together) or was a feature. For example, if the intended utterance were Dan's cap, feature errors might be tan's gap or gan's tap, but a whole segment error would be can's dap. Of the 70 errors in their corpus, 67 were whole segment errors, apparently strongly favoring segments as the more likely error unit.

There are several complexities in interpretation of these findings, however. The most important one here is that Shattuck-Hufnagel and Klatt (1979) considered only the possibilities that errors involved features or segments. They did not consider other options (cf. Browman and Goldstein 1990), such as syllable constituents. Nor did they provide their corpus of 70 errors so that other options could be assessed. However, it is likely, given their high frequency among sublexical errors, that most of the 70 items involved not only consonants, but also syllable onsets that were also word onsets.

Research beginning with Mowrey and MacKay (1990) has brought errors research into the physiology laboratory where errorful articulatory actions can be observed. Thanks largely to more recent research by Pouplier and colleagues (e.g., Pouplier 2003; Goldstein et al. 2007), we now know that errors are characteristically gradient, not all-or-none, in magnitude and, in contrast to conclusions based on transcription-based errors, are frequently phonotactically illegal. Many errors that are observed in the laboratory are inaudible, and so counterparts are not found in transcription-based corpora.

One important finding from this research domain (e.g., Goldstein et al. 2007) is that single-gesture/feature errors do occur with some frequency in contrast to
observations made based on transcription of errors heard outside the laboratory. For example, in alternating productions of such word pairs as *bad bang*, errors occur in which, for example, an inappropriate velum lowering gesture occurs during intended *bad* without the associated tongue dorsum gesture of *ng* or vice versa. Even so, Goldstein *et al.* (2007) found that, overall, intrusions of both gestures of a nasal consonant occurred together with greater than chance frequency, suggesting a coupling between the component gestures of a nasal.

In short, errors collected in the laboratory provide evidence supporting the occurrence both of erroneous intrusions of single gestures of consonants composed of two or more gestures and of the component gestures of a two‐gesture segment. That is, findings from Goldstein *et al.* (2007) do hint at coupling among gestures to form conventional segments. More research of errors involving such multi‐gesture segments will be required before any firm conclusions are warranted about the reality of conventional segments from this evidence source.

9 The alphabetic principle

A number of researchers and theorists (e.g., Cowley 2011; Port 2010) have suggested that our ideas about the units of the spoken language are strongly influenced, and perhaps are misled by, the units of alphabetic writing systems. Certainly, it is true that metalinguistic awareness of the segmental structure of language is elusive to preliterate individuals (e.g. Liberman 1973) and appears to develop only with the development of literacy in an alphabetic writing system (e.g., Morais, Alegria, and Content 1987). However, before we conclude that phonological segments are fictional reflections of the way that some languages are written, we need to address two issues.

One concerns the basis on which letters were selected as representations of the spoken language when alphabetic systems were invented. The other concerns the, to me, astonishing effectiveness of alphabetic writing systems in making reading possible and even easy to do.

Nonalphabetic writing systems include logographic systems in which written forms map to words or morphemes and syllabic systems in which written forms map to syllables or morae. In all of these instances, the written forms map to presumed real, identifiable units in the spoken language. Is it plausible to suggest that the letters of alphabetic writing systems map to fictional units of the spoken language? My answer is mostly no. In my view, the language‐form particles that underlie duality of patterning in language must be *roughly* segment‐like in size to make sense of the invention of alphabetic writing systems.

Relatedly, written language is parasitic on the spoken language, and reading is parasitic on speech (e.g., Mattingly and Kavanagh 1972). That is, most of us read the very language we have earlier learned to speak, and even very skilled readers access phonology from print (Frost 1998). However, even though the spoken language is primary (in coming first in our species and in ontogeny), it is remarkable the extent
to which we can become proficient at reading and writing. Many of us who have taught phonetics to undergraduates are aware of the difficulty of teaching them that phonetics is about spoken language forms not about letters. Why, if letters of the alphabet map to fictional units of the spoken language can reading and writing become as effortless as they can become and can take hold in the way that they take hold?

10  Languages are not tidy

That asked, it should be acknowledged that languages are not tidy. They consist of conventional patterns that emerge in the context of communicative and social interactions among members of a community (e.g., Millikan 2003). Linguistic description provides an idealization of the nature of those conventions. In the wild, language use is not so tidy as those descriptions might suggest. Language is always changing, and different speakers of the "same" language use it in somewhat different ways. Duality of patterning may be possible, speech errors may look somewhat segmental, alphabetic writing systems may be highly learnable and usable without segments being exactly, wholly, exclusively, units of the spoken language. They may be approximately units of the language, units in the limit.

11  Coupling relations among gestures in articulatory phonology

In articulatory phonology, it is not typically shown in gestural scores that some component gestures of a word are coupled. Although nothing in a gestural score of palm suggests that the labial closure gesture and devoicing gestures of /p/ “belong” together any more than either belongs with the vocalic gesture with which each also overlaps in time, there are differences in how the gestures are coupled. These differences may underlie findings by Goldstein et al. (2007) described earlier in which gestures of nasals tend to intrude together with greater than chance likelihood. Here we assess the extent to which conventional segments emerge from coupling relations among gestures in articulatory phonology.

For movements to be effective, they must be coordinated, and coordination means that there are appropriate coupling relations among bodily segments. For oscillatory actions such as the oscillation of fish fins, or the cycling of human limbs, coupled body segments (fins, legs) tend to move in systematic phase relations. For example, when humans walk, the legs cycle 180 degrees out of phase (with a whole cycle being 360 degrees), that is, in an anti-phase mode in which one leg is one half cycle out of phase with respect to the other. Across varieties of oscillatory actions, stable relations that are frequently observed are in-phase (0 degree phase relation) and anti-phase modes, with the former typically more stable than the latter (e.g., Kelso 1995). Although
other stable modes can be learned, these two modes appear to be intrinsically stable, that is, stable without having to be practiced.

Goldstein, Byrd, and Saltzman (2006) suggest that phonetic gestures also exhibit either in-phase or anti-phase modes. Previous research suggested that consonantal gestures in syllable onsets are phased differently with respect to the vowel than are consonantal gestures in syllable codas. Specifically, onset consonants exhibit a “C-center” timing relation to the vowel (e.g., Browman and Goldstein 2000) so that the temporal midpoint of the consonants in an onset exhibit an approximately invariant timing relation to a following vowel. In contrast, the first consonant gesture in a coda is phased invariantly with respect to the vowel, and any subsequent consonant gestures in the coda are phased with respect to a preceding gesture. Based on these findings, Goldstein et al. (2006) propose that onset and coda gestures exhibit different kinds of couplings with respect to a syllable’s vowel. They propose that oral constriction gestures in the syllable onset are coupled in-phase to the vocalic gesture. By itself, this would yield consonant gestures that overlapped one another completely in syllable onsets. To prevent that, and hence to ensure recoverability of the gestures by listeners, oral consonant gestures in a syllable onset are coupled anti-phase to one to the other. The outcome of the competition between in-phase coupling to the vowel and anti-phase coupling within the onset is the observed C-center relation of onset consonants to the vowel. Because in-phase oral constriction gestures and devoicing gestures or velum lowering gestures would not affect gesture recoverability and in-phase coupling is the more stable coupling, devoicing and velum gestures are coupled in-phase with oral constriction gestures in a syllable onset.

Figure 2.4 shows an example of coupling relations of consonant gestures in an onset and vowel for the word spot as depicted by Goldstein et al. (2006; their Figure 7.8). In the bottom half of the figure, the gestural score for spot is shown with coupling relations among gestures overlaid. The set of coupling relations (the “coupling graph”) is also projected for easier viewing in the upper portion of the figure. In either display of the coupling relations, solid lines represent in-phase coupling modes; dotted lines represent anti-phase coupling modes. Accordingly, the alveolar constriction gesture for /s/ and the labial closure gesture for /p/ are both coupled in-phase with the vowel (the pharyngeal gesture in the figure), but anti-phase with one another. The devoicing gesture shared by /s/ and /p/ is shown coupled in-phase with the alveolar gesture for /s/. Figure 2.4 shows that the first consonant gesture in the syllable coda is phased anti-phase to the vowel gesture. Not shown is that subsequent coda gestures, if any, are coupled anti-phase with the preceding gesture.

For our example word palm, the two onset gestures (for /p/) would be coupled in-phase to the vowel and to each other; the two coda gestures (for /m/) would be coupled anti-phase to one another, and one of them (the velum lowering gesture) would be coupled anti-phase to the preceding vowel. Nothing in this set of coupling relations suggest that the two gestures of the conventional segment /p/ or of the segment /m/ “belong” to one another in any special way. That is, segments do not appear to emerge as special units in this augmentation of the gestural score.
Another dimension of gestural relations that might foster emergence of segmental structure from gestural constellations is coupling strength (cf. “bonding strength,” Browman and Goldstein 2000). Are the gestures that compose conventional segments (so, the devoicing and oral constriction gestures of voiceless obstruents, the velum lowering and oral constriction gestures of nasal consonants) more tightly coupled one to the other than each is to gestures for other neighboring gestures in a gestural constellation?

The data here are sparse. However, Munhall, Löfqvist, and Kelso (1994) offer positive evidence. In a study of effects of lip perturbation on gestural timing (cf. Kelso et al. 1984 discussed above), Munhall et al. used a lip paddle to delay the lower lips’ contribution to lip closure for the first /p/ in /pip/. The perturbation had the effect of delaying laryngeal abduction, the other gesture of /p/, but did not delay the following vocalic gesture. This might index a tighter coupling between the two gestures of /p/ than between the oral constriction gestures of /p/ and /i/. More findings like that might show that gestures composing a conventional segment show especially strong coupling relations.

13 Hidden gestures?

There is a way to address another problem for the idea that conventional segments are units for talkers that Figure 2.3 revealed. In the words sod and pod, there are devoicing gestures associated with the initial consonants /s/ and /p/. However, in spod, there is just
one devoicing gesture. To which segment does it belong? In a conventional analysis, /s/ and /p/ are separate segments (although there are proposals (e.g., Fudge 1969) that they jointly constitute a single complex segment). If so, each should have its own coupled devoicing gesture. A somewhat unappealing way to preserve the notion of segment in this instance is offered by another finding by Munhall and Löfqvist (1992). Figure 2.5 shows three panels (selected from their Figure 4) of a single speaker producing the utterance Kiss Ted at three speaking rates. The figure shows waveforms below tracings of glottal opening and closing gestures. At the slowest rate, there are distinct peaks of laryngeal opening for the /k/ and /s/ in kiss and the /t/ in Ted. At the intermediate rate, the gestures for /s/ and /t/ have begun to merge; at the fastest rate, just one gesture is visible for the two consonants. Perhaps in the last instance, and in the /sp/ cluster of spod, there are two devoicing gestures “underlyingly” that, however, overlap completely.

However, this account may be too much of a reach for a phonology such as articulatory phonology that is meant to stay close to the articulatory surface. In any case, there are yet other indications that gestures do not always coalesce into conventional segments, at least not in a transparent way. The component gestures of multi-gesture segments, including nasals (Krakow 1989), /l/ (Sproat and Fujimura 1993), /r/ (Gick and Campbell 2003) and /w/ (Gick 2003) are phased according to the description provided earlier for onset and coda consonants, and therefore are phased differently in onsets and codas. Across a variety of languages, the gestures for these multi-gesture segments are produced synchronously before the vowel but are phased according to a sonority-like hierarchy postvocically (Gick, Campbell, Oh, and Tamburri-Watt 2006). Why do /m/s, /l/s, /r/s, and /w/s before and after vowels count as instances of the same segment?
14 Concluding remarks

In my opinion, articulatory phonology offers the most successful effort to date to develop a realistic perspective on the primitives of language forms. Its gestural primitives are realistic in the important sense that they are designed for public language use.3

Within that perspective the answer to whether there are segments in the conventional sense, that is, that correspond to segments as transcribed in IPA notation appears to be “only almost.” A naysayer might even say “nay.”

There have to be particles of form that underlie duality of patterning and the particulate principle. Meaningless particles of form (or almost meaningless particles; see, e.g., Kunihira 1971; Perniss, Thompson, and Vigliocco 2010; Remez, Fellowes, Blumenthal, and Nagel 2003) have the property of being able to be combined in new ways without blending to generate new word forms. But, of course, the particles do not have to be conventional segments. There is no convincing evidence from any source, to my knowledge, that they are. However, I am convinced by the success of alphabetic writing systems, and the approximately segmental character of a substantial subset of sublexical speech errors that the particles are not far from conventional segments. Further research on speech errors elicited in the laboratory and on gestural coupling strength is required to test this conclusion further, however.

An important observation from the varieties of data I reviewed in a search for segments from an articulatory phonological perspective, is that languages are not tidy. As emergent, ever-changing systems that are known somewhat differently by different users of the “same” language, languages are only as tidy (in having particles of language form of a particular kind, in permitting clear classification of forms into segments or not, and, for that matter, words or not) as they have to be to do their work in the world. There are only almost segments.

Acknowledgments

Preparation of the manuscript was supported by grant HD-001994 to Haskins Laboratories. I thank Louis Goldstein for his comments on a draft of this chapter.

Notes

1 I am not intending to take a stand here on the issue whether human language originated as a gestural system (e.g., Hewes 1973; Corballis 1991; but see MacNeilage 2008). Either way, today spoken language is universal.

2 “relative” time, because an entry should be abstracted away from time, say, in milliseconds.

3 One can ask why acoustic primitives might not also be realistic especially from the perspective of the listener. For reasons why they would not be, see Fowler (1986, 1996).
References


Beyond the Segment

Markus A. Pöchtrager

1 Introduction

A segment is the smallest unit we arrive at by cutting the signal vertically, for example a word like mass is usually cut into [m] + [æ] + [s]. The phoneme is nothing but a special type of segment, one that is stripped of any non-distinctive information, namely of anything that does not serve to differentiate meaning. So much for the conventional wisdom found in any introductory textbook to phonology. How we arrive at the vertical cuts that delineate segments is of course not a trivial question, as sound waves obviously do not have little breaks between individual sounds. Whether the sound before the vowel in the word chess counts as one or two segments is a decision that must be informed by phonological theory. Despite such difficulties, the segment has held pride of place in most phonological theories, but as Anderson (1974: 6) points out, “[t]he only justification that can be given is a pragmatic one; such a description, based on a segmental structure imposed on the event by the analyst, has been the basis of virtually every result of note that has ever been obtained in the field of linguistic phonetics or phonology.” And, as Anderson (1974) goes on to say, the segment “will continue to be employed until some deficiency is pointed out, in the form of a linguistically significant generalization that is essentially unstatable if the procedure of segmentation is adhered to.”

It is clear that one such deficiency was pointed out by Autosegmental Phonology (Goldsmith 1976), which took the rejection of the “Absolute Slicing Hypothesis” as its starting point. That hypothesis states that a phonological string can be cut up into a series of fine slices (“segments”), each of which is characterized by a number of properties, and where the properties of one slice are in theory completely independent of the properties of the adjacent slices. This is of course the view of SPE (Chomsky...
and Halle 1968), with its feature matrices, and later models that derive from it. Goldsmith rejected that view, pointing out that there are many phonological properties, for example tonal features, that extend over more than just one segment; hence the notion “autosegment” for a unit that is independent of the segment. It is worth noting, however, that even in Autosegmental Phonology the notion of segment does not disappear altogether, but is only downplayed. Phonological properties might not be unique to one segment (but shared), but the segment still acts as a kind of reference point, otherwise we could not express the extension of a given property. Furthermore, ideas about how to segment a given word are often a leftover from earlier phonemic approaches: in English *sent*, coronality might be shared between the *n* and the *t*, but still there are four segments (as there were four phonemes) in most analyses.

In this chapter I want to address yet another deficiency of the notion of segment: too often, the segment leads us to believe that it is the locus of independent phonological difference; that the difference between two words has to be located in one single place only. Against this view, I will argue that what characterizes the difference between words can be a larger pattern, in particular a trade-off between vowel length and consonant length. Since in such trade-off patterns both the vowel and the consonant differ at the same time, it is a moot point whether one or the other only is responsible for differentiating words.

Before we proceed, there is one issue that needs clarification. There is a logical relationship as expressed under (1), such that if one believes in phonemes, one necessarily has to believe in segments, and if one believes in segments, one has to believe in discreteness of some sort.

(1) (a) phoneme → (b) segment → (c) discreteness

The reverse, however, does not hold. While the view on phonological representations presented here subscribes to discrete units (1c), though relatively fine-grained ones, it does not follow that the segment or the phoneme as traditionally understood (1a–b) play any role. As we shall see, the traditional notion of segment is in fact an impediment to the understanding of certain phonological phenomena; in our case, length in English and Estonian. In those two languages, it is the trade-off between vowel length and consonant length that is important: the length of the vowel and that of the consonant stand in an inverse relationship with each other. The notion of “phoneme” (a “segment” that, in addition, differentiates meaning) is equally cumbersome, as it brings with it the division between phonemic and allophonic, which will make it impossible to see the clear parallels that exist between Estonian and English.¹

In the course of dealing with those parallels, we will make use of a particular representational framework as developed in Pöchtrager (2006). Note that the thrust of the argument does not rest on the particular representational implementation chosen. Whichever model is adopted, it should be able to deal with the trade-offs to be discussed.
2 Binary and ternary systems of length

Estonian is famous for its allegedly unusual length system and has puzzled linguists for a long time (cf., amongst others, Bye 1997; Hint 1973, 1998; Lehiste 1960, 1965; Ojamaa 1976; Posti 1950; Prince 1980; Tauli 1973 etc.). In order to understand the consternation and to put things into perspective, let us compare Estonian (4) to English (2) and Italian (3). While English or Italian make do with a distinction between short and long, Estonian displays three degrees of length (short, long, overlong). Note that the charts in (2–4) give “accepted transcriptions” which, as we shall see, are actually misleading.2 ([ː] marks length and [ːː] overlength.)

(2) English:  
bit [bɪt] ≠ beat [biːt]  
bid [bɪd] ≠ bead [bɪːd]  
full [fʊl] ≠ fool [fuːl]  
bet [bet] ≠ bait [beɪt]

(3) Italian:  
fato ['faːto] ‘fate’  
casa ['kaːza] ‘house’  
V V C
fatto ['fatːo] ‘done’  
cassa ['kasːa] ‘till’  
V C C

(4) Estonian:  
a. [lina] ‘linen (nom.sg.)’  
 [sada] ‘hundred (nom.sg.)’

b. [linːa] ‘city (gen.sg.)’  
 [saːda] ‘send! (imper.)’

c. [linːːa] ‘city (par.sg.)’  
 [saːːda] ‘to receive (inf.)’

Several comments are in order. First, the choice of a long (beat, bead) or a short vowel (bit, bid) in English in (2) seems unaffected by the choice of the final consonant –“seems,” because we will see in a moment that this is actually incorrect; it is an artifact of the “accepted transcription.” (Note also that some phonological interpretations choose to ignore the vowel length in bit vs. beat and claim that it is the laxness/tenseness that counts. For arguments that both properties are needed, compare Kaye 1995: 318ff.) Second, Italian displays an interesting trade-off in length: the more room taken up by the consonant (C), the less remains for the preceding vowel (V) and vice versa. (In (3), this is symbolized by an asymmetric distribution of Vs and Cs to the right of the chart.) Third, bisyllabic words such as the ones given in (4) are claimed to establish a three-way segmental (“phonemic”) contrast for Estonian. That three-way distinction makes Estonian look very different from the other languages.

Before we conclude that Estonian is simply “unusual,” we need to consider a couple of facts about Estonian that often go unnoticed or are chosen to be ignored. In fact, many of the problems traditional accounts of Estonian struggle with are simply an artifact of the particular framework adopted in those analyses. Consider the chart in (5) which gives monosyllabic words (in contrast to the bisyllabic words in (4)); the words on the left are taken from Ojamaa (1976: 9 and passim).
(5) Estonian:

a. [geː:b] ‘it boils’  [siː:d] ‘silk (nom.sg.)’  V V V C
b. [geː:b] ‘cape (nom.sg.)’  [giː:dː] ‘praise (nom.sg.)’  V V C C
c. [gebː] ‘stick (nom.sg.)’  [judː] ‘story (nom.sg.)’  V C C C

The monosyllabic words in (5) show a trade-off similar to Italian: the more room taken up by the consonant (C), the less remains for the preceding vowel (V) and vice versa. Such trade-offs show that phonological differences can have multiple exposition, unlike what the notion of segment suggests: the difference between (5a) and (5b) lies not only in the length of the vowel, but also that of the consonant; analogously for (5b) and (5c). The triplets in (5) are of course problematic for a phonemic analysis, as was pointed out as early as Ojamaa (1976). There is a mutual dependency between the vowel and the consonant in monosyllabic words. In a phonemic analysis of the data in (5), either the length of the vowel could be chosen as phonemic (in which case that of the consonant would be allophonic) or the other way round. That is, if we only knew the length of the vowel, we could predict the length of the consonant. Likewise, if we only knew the length of the consonant, we could predict the length of the vowel. Which one is it to be? No conclusive decision either way is possible. If we add to this the bisyllabic words in (3), one would have to conclude that both vowel length and consonant length are phonemic, in which case the trade-off seen in monosyllables (5) becomes even more mysterious. Why should there be a trade-off in (5) if the length of the vowel and the consonant can vary independently as per (3)? The problem with (5) clearly stems from the assumption that phonological differences must be located in one single position only; an assumption which in turn is based on the data in (4). No insightful analysis emerges.

There is yet another problem with (4), and that is that the transcriptions are simplified and adjusted. The chart in (6) repeats that in (4), but with more accurate transcriptions.

(6) Estonian:

a. [liːnaˑ] ‘linen (nom.sg.)’  [saːdaˑ] ‘hundred (nom.sg.)’
b. [linːaˑ] ‘city (gen.sg.)’  [saːdaˑ] ‘send! (imper.)’
c. [linːːa] ‘city (par.sg.)’  [saːːda] ‘to receive (inf.)’

The chart in (4) disregarded differences in the final vowel. As (6) makes clear, it is slightly longer in (6a) and (6b) than in (6c), indicated by [ˑ]. The lesson from (6) is similar to the one from (5). Differences between individual words are not located in one single point. What sets apart (6a), (6b), and (6c) is not only the length of the intervocalic n (left column) or the length of the first a (right column), but also that of the final vowel.

The data in (6) have also provided grist for the mill of those arguing against three phonemic degrees of length. While the monosyllabic words in (5) are not often discussed in the literature on Estonian, those in (6) certainly have been hotly debated.
An analysis with three phonemic degrees of length, as (4) seemed to suggest, sits uneasy with the often implicit, sometimes explicit assumption that languages only ever distinguish between two degrees of length. An early statement to that extent comes from Trubetzkoy (1938), in whose opinion any analysis (of any language) that involves more than two degrees of length is based on a misunderstanding. Accordingly, many analyses have been put forward to make Estonian more like other languages, trying to reduce the ternary distinction to two independent parameters. The earliest example is probably Trubetzkoy (1938) himself. Posti (1950), basing himself on data like those in (6), argues that phonologically, Estonian only distinguishes short and long. For him, these attributes characterize entire syllables and not individual sounds. Where other accounts differentiate between a long and an overlong $\alpha$ (5b–c), for example, he assumes that it is the length difference of the final syllable (realized on the vowel) that is responsible for the meaning difference. The following chart illustrates Posti’s interpretation.

(7) Posti’s phonological interpretation

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[sadaˑ]</td>
<td>short + long</td>
</tr>
<tr>
<td>b.</td>
<td>[saːdaˑ]</td>
<td>long + long</td>
</tr>
<tr>
<td>c.</td>
<td>[saːːda]</td>
<td>long + short</td>
</tr>
</tbody>
</table>

Posti needs to exclude the combination short + short by stipulation, that is, the analysis overgenerates. What is more troubling, of course, is that his analysis cannot be extended to the monosyllabic words in (5). If there is only a difference between short/long syllables, how can there be three different types of monosyllables?

In a similar vein, Tauli (1973) also claims that Estonian only distinguishes between short and long. For him, these are properties of individual vowels and consonants. But unlike Posti, Tauli postulates that it is the property of bearing heavy stress or light stress that leads to an additional distinction. Such heavy/light stress “is a feature which distinguishes two primary- or secondary-stressed syllables with the same phonemic composition from each other” (p. 17). This is illustrated in (8).

(8) Tauli’s phonological interpretation

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>bisyllabic</td>
<td>monosyllabic</td>
<td></td>
</tr>
<tr>
<td>a.</td>
<td>[sadaˑ]</td>
<td>/sata/</td>
</tr>
<tr>
<td>b.</td>
<td>[saːdaˑ]</td>
<td>/saata/</td>
</tr>
<tr>
<td>c.</td>
<td>[saːːda]</td>
<td>/saata/ (heavy stress)</td>
</tr>
<tr>
<td>d.</td>
<td>[geː:b]</td>
<td>/keep/</td>
</tr>
<tr>
<td>e.</td>
<td>[geːbː]</td>
<td>/keepp/</td>
</tr>
<tr>
<td>f.</td>
<td>[gebːː]</td>
<td>/kepp/</td>
</tr>
</tbody>
</table>

(8b) and (8c) are identical except that the former has light, the latter heavy stress. However, Tauli’s analysis (like that of Posti) overgenerates: we should get $2 \times 2 = 4$ logical possibilities, but only three occur, compare (8a–c). Tauli simply stipulates that in a short syllable, that is, one without a long vowel or consonant, “the weight of stress [i.e., the distinction light/heavy stress] is neutralized” (Tauli 1973). Why it is neutralized remains unaddressed. Similarly, in monosyllabic words (8d–e), we are missing the combination short vowel plus short
consonant (an issue unaddressed by Tauli) and weight of stress must be neutralized here, too, likewise unexplained.

Note that there is yet another problem with the analyses in the wake of Trubetzkoy: the discussion revolves around the correct way to interpret Estonian length differences. Posti or Tauli do not deny that the nasal in [linːːa] in (6c) is measurably longer than that of [linːa] in (6b), but they argue that that difference is phonologically irrelevant, because it is predictable from other factors. Phonologically, the intervocalic nasal in (6b) and (6c) is identical for them. Yet, in actual pronunciation, the lengths of the n in those two words do differ, and they differ in such a way that the n is shorter if the final vowel is longer and vice versa. In other words, the trade-off does not go away in Posti's analysis, but is simply demoted to a status of “allophonic only.” Why the trade-off should exist in the first place and why it takes the shape it takes remains unexplained.

In the remainder of this chapter we will see that the three degrees of length do not have to be explained away. They are neither exotic nor problematic, but rather more common than usually assumed. In fact, despite first impressions (and misleading hidden assumptions), English and Estonian are in large parts identical. In order to see that, we will need some theoretical armament. We will look at two of the most fundamental principles of Government Phonology (the framework in which the present chapter is couched) and a seemingly wrong prediction that one of them makes.

### 3 English paves the way for Estonian

Government Phonology (GP) has been constructed as a highly restricted theory, that is, it sets out to define exactly what can be a phonological phenomenon and what cannot. As such, it claims a rich empirical content. This is achieved by two basic principles: the Non‐Arbitrariness Principle and the Minimality Hypothesis.

The Non‐Arbitrariness Principle (NAP) is without doubt what makes GP different from any other theory on the “market”:

(9) Non‐Arbitrariness Principle (NAP)

There is a direct relation between a phonological process and the context in which it occurs. (Kaye, Lowenstamm, and Vergnaud 1990: 194)

This imposes a very strict restriction on what phonology can and cannot do. Vowel harmony, for example, involves the adjustment of vowels to neighboring vowels with respect to some property (say, frontness), and, as such, shows a connection between what happens (adjustment of a vowel) and where it happens (in the context of exactly the property that is being adjusted). Velar softening in English (electric vs. electricity), on the other hand, could not be a phonological process, as there is nothing in the theory (of the internal structure of sounds, say) that leads one to expect that k stands in some relationship to s that is somehow connected to i.
The NAP operates in tandem with the Minimality Hypothesis (MH).

(10) Minimality Hypothesis (MH)
    Processes apply whenever their conditions are met. (Kaye 1992: 141)

This excludes several things at the same time. First, the MH excludes exceptions: If a phonological process can take place, it must take place. There is no such thing as “marking as an exception,” and so on. Second, it also excludes extrinsic ordering of processes (cf. rule ordering in most rule-based approaches). A process P cannot be held at bay simply to make sure that another process Q applies before it. If the conditions for P are met, it must take place. Third, the MH also excludes “non-derived environment blocking” (Kiparsky 1976). If a non-derived environment fulfills the required condition, the process will take place and cannot be blocked.

Note that the two principles apply in tandem. The NAP and MH have to be met at the same time in order for a process to be called phonological. To come back to the aforementioned examples, vowel harmony is usually non-arbitrary and exceptionless. Velar softening is neither non-arbitrary nor exceptionless: English has no problem in general with k before i (kicking does not become kissing).

With the restrictions imposed by the NAP and the MH in mind, consider now the chart in (11). The data come from New York City (NYC) English, but the phenomenon it illustrates can be found in many other varieties of English as well.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>bid</td>
<td>bːd</td>
<td></td>
<td>bit</td>
<td>bit</td>
</tr>
<tr>
<td>bead</td>
<td>bːd</td>
<td></td>
<td>beat</td>
<td>biːt</td>
</tr>
<tr>
<td>big</td>
<td>bːg</td>
<td></td>
<td>sick</td>
<td>sik</td>
</tr>
<tr>
<td>league</td>
<td>liːːg</td>
<td></td>
<td>beak</td>
<td>biːk</td>
</tr>
<tr>
<td>rib</td>
<td>nːb</td>
<td></td>
<td>rip</td>
<td>rːp</td>
</tr>
<tr>
<td>lube</td>
<td>luːːb</td>
<td></td>
<td>loop</td>
<td>luːːp</td>
</tr>
<tr>
<td>bin</td>
<td>bːn</td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>bean</td>
<td>bːn</td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>dim</td>
<td>dːm</td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>deem</td>
<td>dːːm</td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>bill</td>
<td>bːl</td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>peel</td>
<td>piːːl</td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>live</td>
<td>liːːv</td>
<td></td>
<td>stiff</td>
<td>stːf</td>
</tr>
<tr>
<td>leave</td>
<td>liːːv</td>
<td></td>
<td>leaf</td>
<td>liːːf</td>
</tr>
<tr>
<td>his</td>
<td>hːːz</td>
<td></td>
<td>hiss</td>
<td>hːːs</td>
</tr>
<tr>
<td>(to) use</td>
<td>juːːz</td>
<td></td>
<td>(a) use</td>
<td>juːːs</td>
</tr>
</tbody>
</table>

The chart contains only monosyllabic words. The words in the column on the left all end in a neutral (traditionally: “voiced”) consonant, those on the right in a voiceless consonant. The vowel is systematically longer in the words on the left, namely, there is a correlation of vowel length and the nature of the following consonant. This phenomenon is well-known in the literature, but usually disregarded as phonologically irrelevant because of its predictability, compare e.g., Clark and Yallop (1995: 33f).
How are we to deal with this? All consonants in the column on the right contain the element \( H \), which delineates the natural class of voiceless sounds. Now, we could express our observation in a statement like the one in (12), which is of course only one way of framing things.

(12) Additional length occurs in monosyllabic words if the vowel is not immediately followed by a phonological expression containing \( H \).

While (12) is certainly a correct observation about the facts, it is far from an explanation of what is going on. It fails the NAP for at least three reasons. First, a melodic property (the element \( H \)) seems to interact with length, which is a structural property (i.e., how much room is taken up). Second, there is no relation between the absence of \( H \) and (additional) length. Why should those two properties be connected in any way? Third, what makes \( H \) different from other melodic properties? Certainly we do not expect to find a language where a vowel is longer before labials (characterized by the element \( U \)).

While (12) is a blatant violation of the NAP, it does conform to the MH. All monosyllabic words of NYC English behave like the ones in (11), without any exception. This is exactly what we would expect of a phonological phenomenon, yet it cannot be expressed in the language of GP. Something will have to give way. In the next section, I will argue that it is the element \( H \) as a melodic property that has to go.

4 The fortis/lenis hypothesis

How to get out of this predicament? The proposal put forth in Pöchtrager (2006) is that \( H \) is not an element, but a particular structural configuration: \( H \) is a type of length. Objects that were thought to contain \( H \) are simply the longer versions of objects without that element. Under such a reinterpretation, the need for an element \( H \) disappears: An English \( d \), for example, is nothing but the “short version” of a \( t \). Otherwise, they are identical. The same holds true for pairs like \( b/p, v/f \), and so on. I will refer to \( b, d, g, v \) etc. as “lenis” or “Q1” (“quantity 1”), and to \( p, t, k, f \) etc. as “fortis” or “Q2.”

With this slight change in our approach to English consonants, a non-arbitrary solution comes into reach. The extra length that we find in English monosyllables like \textit{bid}, \textit{give}, and so on can now be seen for what it is: a trade-off like the one in Italian. The less room taken up by the final consonant, the more remains for the preceding vowel, and so on. (13) giving a schematic expression to this idea.\(^9\)

(13) \begin{tabular}{l} \textit{bid} \quad [bıːd] \quad \textit{give} \quad [gıːv] \quad \textit{V V C} \\ \textit{bit} \quad [bıt] \quad \textit{riff} \quad [rif] \quad \textit{V C C} \end{tabular}

Note that the proposal to reinterpret the difference neutral/voiceless as one of length is not just a trick to get us out of difficult situation: An English \( t \) is measurably longer than a \( d \) (Lisker 1957; Lisker and Abrahamson 1964; Luce and Charles-Luce 1985). What we are saying is that this difference in length is exactly what characterizes the phonological difference between \( d/t \), and so on.
It is now time to formalize the proposal. Since the discussion will be somewhat technical, it is best to start with an example. (14) gives the relevant parts of the pair give and riff, that is, the vowel followed by the final fricative.

(14)  a.  

\[
\begin{array}{c}
\text{xN}\{\text{I}\} \\
\downarrow \\
\text{x}_1 \\
\downarrow \\
\text{xO}\{\text{U}\} \\
\end{array}
\]

b.  

\[
\begin{array}{c}
\text{xN}\{\text{I}\} \\
\downarrow \\
\text{x}_1 \\
\downarrow \\
\text{xO}\{\text{U}\} \\
\end{array}
\]

(14a–b) are identical in terms of positions and constituency. We have three skeletal positions: a nuclear head xN, the position x₁, and an onset head xO. (For arguments that word-final consonants are really onsets compare Kaye 1990, Harris and Gussmann 1998.) The nuclear head xN is annotated with the element I, which defines the quality of the vowel i. The onset xO is annotated with the element U, making it labial. The final onset head xO and the preceding position x₁ form a constituent together, O’ (a projection of xO). The highest projection of xO, O’, dominates two positions, which means we are dealing with a fricative.¹⁰ This O’ in turn is part of N’, the constituent projected up from xN.

What is crucial is the position x₁. Depending on who it is taken by, we get the difference between give or riff. In riff (14b) x₁ is claimed by xO (indicated by an arrow). That is to say, the fricative f uses up both positions, and the preceding vowel must be short. In give (14a), on the other hand, x₁ is not claimed by xO. The fricative v consists of one position only, xO. As a consequence, x₁ can be taken by xN. The vowel now consists of two positions (xN and x₁) and this explains why we have a lengthened vowel before a lenis consonant like v.

Being fortis/lenis is obviously a lexical property of English consonants. The relationship between xO and x₁ is either there or it is not. As we shall see in section 6, the length of vowels is changeable. Whether that means that the relationship indicating a lengthened vowel (between xN and x₁ in (14a)) is only introduced in the course of a derivation (and thus not underlying) or removed in the course of a derivation (and thus potentially underlying), is irrelevant to the present discussion.

The foregoing discussion makes clear how the fortis/lenis hypothesis solves our problems with the NAP: Lenis consonants contain an unused point, which can be taken by the preceding vowel. This is why we get lengthening of the vowel. At the same time, the special status of the old H element becomes clear: in reality, there is no element H; what we are dealing with is a structural configuration.¹¹

The implications are clear: first, the representations of sounds that so far have been thought to take up one position only will contain more than just one point. Phonological representations are more fine-grained than commonly assumed. Second, not only the points and their grouping into constituents matter, but also who takes up how much room: in (14), both words end in an O’ comprising two points.
The difference between v/f only depends on how many positions are claimed by xO. The constituent O’ is in turn part of a larger constituent. The sounds v/f are an epiphenomenon resulting from several factors distributed over a chunk of structure.

To finish the discussion, let us look at (the relevant parts of) the representations of leave, leaf, and bee.

(15) a. leave
   b. leaf
   c. bee

Leave differs from give (14a) in that the vowel is lexically long (expressed by x₂) and can take the additional point contained in v (x₃), giving us an overlong vowel. In leaf x₃ is used up by the f, hence the vowel is only long. In bee there is no final consonant, yet the vowel length is that of leave. This seems to have to do with minimal size requirements and will not be pursued further here. Note that x₃ sits in the position occupied by the final consonant in (15a–b). For further discussion compare Pöchtrager (2006).

5 Estonian meets English

The representational format introduced in the previous section will, as a by-product, also make it possible to understand Estonian. The chart in (16) compares monosyllabic words in both languages, showing that they are more similar than usually assumed.¹²

(16) VOWEL   CONSONANT
   a. —   Q2   Q1   bid   biːd
   —   Q1   Q2   bit   bit
   b. maa maːː ‘country’ Q3 —   bee   biːː
   c. siid siːd ‘silk’ Q3 Q1   bead biːːd
   kiit giːd: ‘praise’ Q2 Q2   beat biːt
   jutt judː: ‘story’ Q1 Q3 —
Consider first (16a). bid/bit are too short for Estonian. This is possibly connected to the English tense/lax system, that is, the fact that in English the vowels in bid/bit are lax, while Estonian vowels are always tense. We will not pursue this any further here; it is clear that this is a point of contrast between the two languages. Estonian is more rigid in what characterizes a minimal word.

Let us move on to (16b). Vowel-final monosyllables (rather: words ending in a stressed vowel) are overlong in both languages. The minimal size requirements operative in English, alluded to before, are also at work in Estonian.

Finally, we come to (16c). What characterizes the words in (16c) is that a total of four positions can be divided between the vowel and the final consonant. This gives us three logical possibilities: Q3+Q1, Q2+Q2 or Q1+Q3. Estonian makes use of all three. This is of course the pattern we had seen at the beginning of the chapter, in (6). English is more modest; the last pattern is not attested. This is another way of saying that English has no overlong consonants.

The upshot of (16) is that half of the forms have identical representations in both languages. Note that even just stating such a parallel would have been impossible if we had stuck to the difference between phonemic and allophonic.

Let us take stock: first, looking at a property that is considered irrelevant (“allophonic”) in English, viz. additional vowel length before lenis consonants, has opened the door to a representational format that can also handle the Estonian length system (“phonemic”). The distinction between phonemic and allophonic clouds our view, making it impossible to see the clear parallels between Estonian and English (in addition to the problems discussed in section 2). And still, any difference that does exist between the two languages can be expressed by other (independently necessary) means, such as what the minimal size of words is. Second, the notion of trade-off is more important than an individual segment. The difference between two forms is located in the vowel and the consonant, and not just in one or the other (contra the traditional segmental view).

6 Bisyllabic words

The parallels between Estonian and English do not end here. So far we had looked at monosyllabic words where the position not used by a final lenis consonant could be used by the preceding vowel. This position is not available if the consonant in question belongs to the next “syllable”; compare these English data:\textsuperscript{13}

\begin{tabular}{llll}
(17) & a. rub & rʌ:b & rubber & ˈrʌbә \\
 & b. men & me:n & many & ˈmeni \\
 & c. leave & liːv & beaver & ˈbeәvә \\
\end{tabular}

The \textit{b} in \textit{rub} and \textit{rubber} is lenis, yet while we get a lengthened vowel in \textit{rub}, we do not in \textit{rubber}; the same holds true for the other cases in (17). This difference
between monosyllabic and bisyllabic words is true irrespective of whether morphology is involved or not:

(18) a. tube tuː:b tuba 'tuːbə' no morphology
    b. soup suːp super 'suːpə' no morphology
    c. lube luː:b lubing 'luːbɪŋ' morphology
    d. loop luː:p looping 'luːpɪŋ' morphology
    e. seed siː:d seeding 'siːrŋ' morphology
    f. seat siːt seating 'siːrŋ' morphology

In monosyllabic pairs such as tube and soup we see a clear length difference in the vowel in addition to the difference in the final consonants. No such length difference in the vowel can be seen in (morphologically simplex) tuba and super, despite the fact that one has b, the other p. The same is true for morphologically complex cases: Compare lube and loop (different in vowel length and final consonant) to lubing and looping (different in stem-final consonant only). This should also lead us to expect that if, for whatever reason, the difference in the stem-final consonants was lost, the two forms of the gerund should sound exactly identical. An example like that is furnished by seed and seat. While the stems by themselves differ in vowel length and final consonant, once the suffix -ing is added, tapping occurs, wiping out the only difference possible in a bisyllabic form, namely, the difference in the consonants.14

What is particularly interesting for us is that we observe a surprisingly similar effect in Estonian. A word like siid [siːːd] “silk” has a Q3 vowel, yet its genitive siidi [siːːdiˑ] has a Q2 vowel in the stem. Like in English, the extra length made possible by the lenis consonant in monosyllabic forms does not occur once the word is bisyllabic. (Note that the final vowel, i.e., that of the suffix, has so-called half-length, a point we will come back to in a moment.) English and Estonian not only proved to be similar in monosyllabic words, but also in bisyllabic cases.

The question will be: how to make sense of this? Why is there a difference between monosyllabic and bisyllabic words? In both languages, the extra room given by a lenis consonant is available unless the vowel of the following syllable takes precedence. Consider the following schematic diagram of what is going on in Estonian:

(19) siːd STEM/NOM.SG. o o ( o • )
     i SUFFIX o
     siːdi∗ GEN.SG. o o ( o • ) o

The (monosyllabic) stem, which is identical to the nominative in this word, has a Q3 vowel. One point of this Q3 vowel is part of the final consonant. In the genitive, a suffix is attached and the form becomes bisyllabic. What is interesting about Estonian
is that the vowel of the suffix snatches the point within the stem-final consonant and becomes longer itself.

While no such lengthening of the final vowel occurs in English (compare happy and abbey), the languages do share an important trait: in bisyllabic words, an unused point contributed by a lenis consonant cannot be used by the preceding vowel. (This is why the vowel in English leave is longer than the one in beaver.) The two languages differ in what happens to that particular point, though. In English, if the unused position in the lenis consonant is not used by the preceding vowel, nobody can use it. In Estonian, on the other hand, the following vowel can use the unused position in the lenis consonant, if the preceding vowel cannot. This will explain the so-called half-long vowel, to which we will return in just a moment. Before that, let us try to understand why the point contributed by a lenis consonant cannot be used by the preceding vowel in a bisyllabic word.

In short, this can be understood as a locality issue. Consider the following abbreviated representations; for details compare Pöchtrager (2006).

(20) 

\[
\begin{array}{c}
V \sim \sigma \\
V \sim \text{foot} \\
\end{array}
\]

In a monosyllabic word like man, the vowel and the final consonant form a constituent. Within this constituent, a trade-off in length can take place: since n is a lenis consonant, it has an unused point, and it is exactly this point that makes the preceding vowel longer. In bisyllabic many, on the other hand, the n forms a constituent with the following nucleus. This constituent is embedded in the constituent headed by the preceding vowel. The consequence is that the lenis consonant n is simply too far away from the preceding nucleus. Distance ("too far") can be computed by constituency; the immediate constituent containing the lenis consonant does not contain the preceding vowel. Accordingly, no interaction can take place.

At the same time, such a representation of bisyllabic words allows us to make sense of the so-called half-long vowel mentioned in section 2. The word siidi [siːdiˑ] “silk gen.sg.” ends in such a half-long vowel. That is, in exactly the words where the first nucleus cannot take advantage of the unused point in a lenis consonant, the second nucleus is longer. This suggests that when the unused point in the lenis d of siidi is not available to the first nucleus, the second nucleus can take it. In fact, this is not surprising given the representation in (20). In a bisyllabic word, any intervocalic
consonant will form a constituent with the following nucleus. That the two should interact is therefore not surprising.

In fact, what we see in the genitive *siidi* is just part of the bigger picture. Consider (21).

**Table 21**

<table>
<thead>
<tr>
<th>Nom.</th>
<th>Gen.</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. <em>siid</em></td>
<td><em>siidi</em> ['si:di:]</td>
<td>‘silk’</td>
</tr>
<tr>
<td>b. <em>kiit</em></td>
<td><em>kiidu</em> ['gi:du:]</td>
<td>‘praise’</td>
</tr>
<tr>
<td>c. <em>jutt</em></td>
<td><em>jutu</em> ['jud:u:]</td>
<td>‘story’</td>
</tr>
</tbody>
</table>

The nominatives are all monosyllabic. They illustrate the three logical possibilities of monosyllabic words that we have seen so far: Q3 vowel plus Q1 consonant, Q2 vowel plus Q2 consonant, and Q1 vowel plus Q3 consonant. Note that in all the words in (21) the distribution of length in the nominative is different from the distribution in the genitive. In the nominative forms we always have a total of four points that have to be distributed between vowel and consonant (Q3 + Q1 or Q2 + Q2 or Q1 + Q3), while in the genitive we only have three (Q2 + Q1 or Q1 + Q2). All the genitives end in a half-long vowel. We have already discussed this for the word for “silk” (21a). Let us now consider the word for “praise” (21b). In its nominative form, it ends in a Q2 consonant. In the genitive form, we find a Q1 consonant in its stead. That is, instead of a fortis consonant we find a lenis consonant, leaving one of its positions unused. That unused position can then be taken by the following vowel, giving us yet again a half-long vowel. We end up with a form that is the same as the genitive in (21a). The last form, “story” in (21c), ends in a Q3 consonant. Again, one of the three positions has to be given up in the genitive, making it accessible to the final vowel. The intervocalic consonant shortens to the “advantage” of the following nucleus. Again, a trade-off can be seen.

In the next section we will address yet another aspect of the length alternations in Estonian. This will show us that the alternations in length are not purely mechanical or inevitable, but rather require reference to morphological domains.

### 7 When morphology intervenes

The previous section discussed similarities between Estonian and English. In this section, for the sake of completeness, we need to address the role played by morphology. At this point, Estonian and English part ways: while the following is crucial to a full understanding of what is going on in Estonian, no counterpart in English can be found. As the chart in (18) showed, what English cares about is whether a form is mono- or bisyllabic; its morphological composition plays no role for length.

Consider the following chart, which is a repetition of (21) with the corresponding forms of the partitive case added on.
Superficially, we seem to have a problem. Like the genitive forms, the partitives in this chart are bisyllabic. Yet the partitive forms do not end in a half-long vowel. Furthermore, no shortening of any kind takes place before them in the stem. In a way, what we want to say is that the partitive is absolutely identical to the nominative with the suffix added on. That is, the nominative siid [siːːd] differs from the partitive siidi [siːːdi] only in that the partitive has a suffix -i, while the nominative does not. The vowel in the stem is overlong in both the nominative and the partitive. The same thing can be said for the other three words in (22). The partitives look just like the nominative plus suffix. This suggests that what we saw in the genitives is not purely mechanical: shortening the stem and creating a final half-long vowel when making the form bisyllabic is not inevitable. It does not happen in the partitive.

How can we understand this difference in behavior between the genitive and the partitive? What I would like to suggest is that we are dealing with a difference in analytic vs. non-analytic morphology (Kaye 1995) here. Let us briefly discuss what this means. Non-analytic morphology is invisible to the phonology; it is treated as if there was no morphology at all. Analytic morphology, on the other hand, is very much visible to the phonology. As an example, take English kept and peeped as discussed in Kaye (1995). From a morphological point of view, both of them must be considered complex; they consist of a stem and a past tense marker. Phonologically, they are quite different, however, in that kept could be a simplex form (cf. apt, adopt etc.), while peeped could never be: no English simplex word could end in a long vowel followed by two stops. This suggests that the two final stops are not in the same domain. That is, in kept we are dealing with non-analytic morphology and the entire string is treated as one single unit by the phonology, [kept]. In peeped we have analytic morphology, [[peep]ed], where phonology applies first to peep, and then again to the entire word peeped. The same difference can be seen in non-analytic parénatal (with stress shift vis-à-vis párënt) and analytic párënhood (where -hood is completely irrelevant to the stress of the base párënt).

Let us apply this to the Estonian words in (22) now. In the genitives, the suffix clearly interacts with the stem. The partitives, on the other hand, look as if the suffix was completely irrelevant to the stem. That is, the genitives involve non-analytic morphology (treated the same as a simplex word), while the partitives are clearly analytic. This is summarized in (23), with square brackets indicating domains.

(23) Nom.Sg. | siːːd  [siːːd]   giː:  [giːːd]   judːː  [judːː]
This squares nicely with one particular fact that is hardly ever stated when talking about Estonian: A phonological string like that represented by the partitives in (23) can only arise through morphological concatenation, in the same way that in English we would never find a long vowel followed by a cluster of two stops, if it were not for analytic morphology. The partitives in (23) give themselves away as morphologically complex by their mere shape. To my knowledge, no other analysis of Estonian even mentions this, they all disregard morphological structure.

If it is true that analytic morphology is a necessary condition for a bisyllabic word to look like one of the partitives in (23), it follows that in a bisyllabic word without morphology we could not possibly have a structure that looks like the partitives in (23). This prediction can readily be tested. Consider the bisyllabic loans from other languages in (24).

(24)  
\begin{align*}
\text{teema} & \quad \text{‘theme’} & [\text{‘deːmaˑ}] \\
\text{floora} & \quad \text{‘flora’} & [\text{‘floːraˑ}] \\
\text{liiga} & \quad \text{‘league’} & [\text{‘liːgaˑ}] \\
\text{loto} & \quad \text{‘lottery’} & [\text{‘lotːoˑ}] \\
\text{loto} & \quad \text{‘lottery’} & [\text{‘lotːoˑ}] \\
\text{summa} & \quad \text{‘sum’} & [\text{‘sumːaˑ}] \\
\text{lasso} & \quad \text{‘lasso’} & [\text{‘lasːoˑ}] \\
\end{align*}

Assuming that loans are treated as single morphemes, we predict Q3 to be impossible. This is indeed correct, as (24) shows. This argument cannot be countered by a claim that the donor languages simply do not have Q3 vowels/consonants. Even if the donor languages really had no Q3 vowels (which I believe to be false), such a counter-argument would run afoul of monosyllabic loans such as the ones in (25), where we do see Q3 vowels and consonants.

(25)  
\begin{align*}
\text{tee} & \quad \text{‘tea’} & [\text{‘deːː}] \\
\text{saal} & \quad \text{‘hall’} & [\text{saːːl}] \\
\text{kool} & \quad \text{‘school’} & [\text{‘goːːl}] \\
\text{loss} & \quad \text{‘castle’} & [\text{‘losːː}] \\
\end{align*}

Clearly then, the lack of Q3 in (24) must be connected to the fact that all those words are bisyllabic and do not involve morphology. This is exactly what we are led to expect by the analysis.

8 Conclusion

The foregoing discussion has tried to show that a lot can be won by giving up certain ideas about phonology in general and English phonology in particular. First, we moved away from the idea that differences between words must always be located in one single point. As we have seen, English bid and bit differ from each other both in their final consonant and the vowel. Mainstream phonological thinking is usually tied to the rather narrow vision of phonology that is claimed to derive from the Saussurean dictum that “[d]ans la langue il n’y a que des différences” (in language there is nothing but differences). Accordingly, a decision has to be made as to what difference is responsible for meaning contrast and every other difference is to be ignored. This is why the vowel length difference in bid/bit is generally neglected,
since it is the difference between $d$ and $t$ that is made responsible for contrast. The implication is that speakers only pay attention to what contrasts. Why this should be so remains unclear. Word-final stops are often unreleased. We live in a noisy world and clearly need every help we can get for parsing an acoustic signal. That the hearer should ignore a reliable clue such as vowel length seems unlikely.

At the same time as we were giving up contrast as a central notion, we gave up the segmental view of phonology that is basically a left-over from phonemics: even in non-linear phonological approaches, one sound is usually equated with one position (except for the obvious case of geminates or so-called contour segments). In the model presented here, on the other hand, one sound usually corresponds to more than one position, and one “sound” can even borrow room from other “sounds,” running against the traditional idea of segments. For English this meant that $d/t$, and so on, do not differ in terms of melody (elements, features), but in terms of length. Phonological representations are more fine-grained than commonly assumed.

While several old-fashioned notions (but nonetheless misleading ones) had to be jettisoned, this shift in perspective made it possible for us to achieve some interesting results. First, it deepened our understanding of the correlation between vowel length and final consonants in English. We could establish a non-arbitrary link between the two. Second, it allowed us to set up a representational format that brings out the commonalities between Estonian (the proverbial “freak of nature”) and English. With that, we have made one further step towards an understanding of Universal Grammar and to what extent languages can differ from each other.

Notes

1 Of course, the phoneme has been under attack for decades, cf. Hamp (1951), Halle (1959), Chomsky (1964), Chomsky and Halle (1968), Postal (1968), Anderson (1974), Sommerstein (1977), to name but a few. But while the segment has been discredited in Autosegmental Phonology, it has never really been removed from its central place in phonological theory. As the discussion of English and Estonian will show yet again, we are better off without it.

2 By “accepted transcriptions” I mean those that can be found in most dictionaries as well as in most phonological treatments of the phenomena in question.

3 Ojamaa (1976: 88) finds that the final vowel in words like (6a) are even longer than in those like (6b). However, the difference is much smaller than that between (6b) and (6c). This does not diminish the thrust of the argument. If anything, Ojamaa’s findings lend further support to the claims of the present chapter.

4 This also helps to avoid circularity: if the theory was restricted by one principle only, say the NAP, we could tweak the theory so as to make something arbitrary into something that is not. However, since the MH has to be satisfied simultaneously, we cannot tamper with the theory so easily. The discussion of English will show us exactly what do to in a situation of conflict between the NAP and the MH. Obviously, a theory constrained by the NAP and the MH will delineate a much narrower field for phonology than do other theories. Most of what counts as phonology in SPE (Chomsky and Halle 1968) and even in many present-day models simply does not qualify as phonology in GP, cf. Kaye (1995) for discussion.
For discussion of the latter point with respect to Turkish vowel harmony cf. Pöchtrager (2010), where it is argued that the so-called exceptions to it are artifacts of a flawed understanding of the Turkish vowel system.

Neutral/voiceless is preferrable to traditional voiced/voiceless, since English b/d/g generally have no vocal fold vibration, as is also recognized in other approaches, e.g., Iverson and Salmons (1995). For recent discussion of the notion “voicing” cf. Harris (2009). The chart in (11) also gives the composition of the individual consonants in terms of elements, e.g., p is \((H, \rho | U)\), i.e., it is a voiceless \((H)\), labial \((U)\) stop \((\rho)\), with \(U\) functioning as the head (underlined). Velars are characterized by an empty head, alveolars by \(A\), and nasality is encoded by \(L\). For further discussion of element theory cf. Charette and Göksel (1994), Harris (1994: Chapter 3), Harris and Lindsey (1995).

While measurements of monosyllabic English words are easily found in the literature (Crystal and House 1988; Denes 1955; Heffner 1937; Klatt 1976; Luce and Charles-Luce 1985; Peterson and Lehiste 1960; Rositzke 1939), measurements for bisyllabic words, discussed in section 6, are harder to come by, but cf. Lehiste (1970), Lisker (1957), Luce, Charles-Luce and McLennan (1999), Umeda (1975). I tried to verify the results found in the literature by doing my own measurements of several sound files in the Oxford Acoustic Phonetic Database (Pickering and Rosner 1993), which, for any given English word, contains 16 recordings (by eight speakers). My findings agree with what the aforementioned sources report.

For arguments against such a position cf. Chomsky (1964: Chapters 4–5) or Harris (1999), who I side with.

Similar claims, i.e., that voiceless consonants are really longer versions of their “voiced” congeners, have been made for other languages, too: Ojibwa (Bloomfield 1956), Cuna (Sherzer 1970), Dutch fricatives (van Oostendorp 2003), Austrian German (Kühnhammer 2004).

Stops involve a second level of projection, i.e., up to \(O^*\), and they comprise a total of three positions. For further details the reader is referred to Pöchtrager (2006).

The implementation of the fortis/lenis hypothesis made several changes to GP’s theory of constituent structure necessary, cf. Pöchtrager (2006). Since the proposal is to give up \(H\) altogether, its other jobs (such as marking high tone) will also have to be taken over by something else. What that “something else” is, is currently unclear.

For the English measurements cf. note 7, for Estonian cf. in particular Ojamaa (1976).

While it is often assumed that the effects of lengthening before a lenis consonant carry over to bisyllabic words (assuming that \(staple/stable\) behave like \(tape/Abe\)), this is in fact somewhat incorrect, cf. the references in note 7. The durational difference of the first vowels in bisyllabic words is extremely small. Umeda (1975: 435) found 10 milliseconds, which I could verify. Lisker (1957: 45) found differences of up to 25 milliseconds, but also states that there are many overlaps. Note that 25 milliseconds is usually considered barely noticeable (Klatt 1976: 1218) and thus any difference found can be considered unreliable and irrelevant.

Zue and Laferriere (1979) claim that there is a length difference in pairs like \(seeding\) and \(seating\), or \(medal\) and \(metal\), for that matter. If anything, their experiments show the opposite, however: participants made an average length difference of 9 milliseconds when \(reading\) pairs like \(seeding\) and \(seating\), but could not tell their own recordings apart in a discriminatory task later on. Nine milliseconds is below the threshold of what can be reliably identified (cf. note 13) and any effect the researchers have found seems to be an artifact of the orthography. For further discussion cf. also Kaye (2012).
This also solves Scott’s (1940) puzzle why Take Gray to London and Take Greater London sound different: a final stressed vowel in English (Gray) will always be Q3, while in bisyllabic Greater we can only have a stressed Q2 vowel.

This is a case of so-called gradation, quite commonplace in Balto-Fennic. While this shortening is phonological, not all cases of gradation are.

Whatever relates keep to kept, they are not related by phonological derivation.

Given that sing-ing has a velar nasal [ŋ] (and not a cluster [ŋɡ]), which is indicative of domain-final position, we must conclude that -ing is analytic (Kaye 1995). But for the distribution of length, English -ing seems to behave very much unlike the analytic suffixes in Estonian, cf. the discussion of tuba and lubing in (18). For further discussion of this rather puzzling matter cf. Pöchtrager (2006).

The analytic/non-analytic distinction suggests in no way that phonology is triggered or blocked by morphological information; the domains in [[peep|ed]] merely show in which order which sub-parts of a phonological representation will be processed. This is very different from, say, Lexical Phonology, where “phonological” rules (e.g., Velar Softening) are not only restricted to certain strata, unlike GP, but also allow for exceptions (Greek/Grec-ist vs. Turk/Turk-ist), unlike GP, and cannot apply to non-derived forms (king), again unlike GP.

References


1 Introduction

Phonetic transcription allows us to put in square brackets many things that languages do not actually make use of, such as palatalized velar glides [ɰʲ] or velarized palatal glides [jˠ]. It also allows us to posit unattested contrasts like pre- vs. postpalatalized sounds [ˈn̥ ~ nʲ] and to entertain what seem to be purely orthographic contrasts like [pja ~ pʲa]. We show here that natural language does not use such refined distinctions and offer a more restrictive theory according to which palatalization, velarization, and labialization are properties only of prosodic levels above the segment.

Following Ladefoged and Maddieson (1996: 2), our study focuses on the elements “that are known to distinguish lexical items within a language,” that is, on minimal-pair contrasts involving labialization, palatalization, and velarization within single morphemes. The facts we present here suggest that natural languages allow at most a single unordered set of vocalic features per syllable margin or word margin, whatever the number of sounds in that domain. For this reason, we propose that the vocalic features behind palatalization, velarization, and labialization occur only once per onset, coda, or prosodic word-edge and are phonologically separate from any consonantal articulations they occur with, so that for example, palatalization occurs only once per onset, regardless of the number of consonantal articulations in that onset. Specifically, we propose that

(1) Each margin of a syllable or word has a single unordered set of vocalic features.
An essential component of our analysis is that the relative timing of secondary and primary articulations \(l^j \text{ vs. } j^l\) is predictable within a language, never contrastive. In this way, (1) models our finding that no language contrastively orders palatalization, velarization, or labialization at any level below the syllable margin.

The idea that vocalic features may characterize prosodic levels above the individual speech sound is not new of course (Harris 1944; Firth 1948, 1957; Goldsmith 1990). What is novel here is our claim that secondary vocalic features only characterize prosodic levels above that level, that no consonant licenses vocalic features on its own. A number of predictions follow from this claim that do not follow from segmental accounts of vocalic licensing. Specifically, we expect to find:

(2) No conflicting vocalic contrasts within a syllable margin
\[ j^l, j^p, p^l, p^j, p^t^l, p^t^j \] are not syllable margins in any language

(3) No pre/post contrasts within a syllable margin
\[ j^p \sim j^p, p^l \sim p^l, j^l \sim j^l, p^t^l \sim p^t^l \] do not contrast in margins in any language
\[ p^l \sim p^l, j^p \sim j^p, p^t^l \sim p^t^l \] do not contrast in margins in any language

(4) No segment/cluster contrasts within a syllable margin
\[ p^l \sim p^l, p^l \sim p^t^l \] do not contrast in margins in any language
\[ p^l \sim j^p, p^l \sim p^j \] do not contrast in margins in any language

We give a few examples of onsets with secondary vocalic features below to illustrate how our prosodic treatment of vocalic features models the restrictions in (2)–(4). We abbreviate the place and manner features of consonants as IPA letters, for example, as \([l]\) or \([pl]\), to focus on our claim that each onset (or coda) has a single set of vocalic features (front, back, round):

\[(5) \text{ Vocalic features are licensed prosodically, not segmentally} \]

\[ \begin{align*}
\text{a.} & \quad \text{Onset} \quad \text{b.} & \quad \text{Onset} \quad \text{c.} & \quad \text{Onset} \\
\text{front} & \quad \text{front} \quad [l] & \quad \text{front} \quad [pl] \\
= [j] & \quad = [^l, j^l, l^j, j^l] & \quad = [^l, p^l, p^l, j^p, p^l, p^j, l^j, j^l, \text{etc.}] \\
\end{align*} \]

(5a) shows a palatal glide, an onset with the vocalic feature \([\text{front}]^2\) and nothing else. Limiting the onset to a single vocalic node, we rule out a velarized \(j\) \([j^l]\), a palatalized \(u\) \([u^l]\), and a palato-velar glide \([ju]\) using phonetically motivated feature co-occurrence restrictions against \([\text{front}]\) and \([\text{back}]\) familiar from previous work (e.g., Hall 1997): nothing can be specified both \([\text{front}]\) and \([\text{back}]\) because those features entail conflicting states of the tongue body.

(5b) shows \([\text{front}]\) linked to an onset with a lateral; we intend no temporal ordering between the feature \([\text{front}]\) and the lateral that is its sister. According to our proposal, (5b) is the phonological representation for both pre- and
A Prosodic Theory of Vocalic Contrasts

67

postpalatalized laterals \([l^\text{i}, \text{i}l]\), which we predict to be allophonic in any language; it also represents lateral-palatal clusters \([l^\text{i}, \text{i}l]\), which we claim are merely orthographic variants of \([l^\text{i}]\) and \([\text{i}l]\), respectively. This models universals (3) and (4) above: no language contrasts pre- and postpalatalized sounds \((l^\text{i} vs. l^\text{i}, \text{i}l vs. \text{i}l)\), or vocalized segments and glide-consonant clusters \((\text{i}l vs. \text{j}l, l^\text{i} vs. \text{j}l)\) we propose, because there is no distinct way of representing them phonologically (see section 4 for further discussion).

(5c) shows palatalization of a \([pl]\) onset: it differs from (5a) and (5b) by addition of additional consonantal articulations, but the same with respect to its vocalic specifications. Given at most one set of vocalic features per margin (1), there is no way to multiply vocalic features in an onset by increasing the number of consonantal articulations there. Thus (5c) represents not only \([p^{l\text{i}}l]\) but also \([pl^\text{i}]\), \([pl^\text{i}]\), and a number of similar things that do not seem to contrast with \([p^{l\text{i}}l]\) in any language. This models (2), (3), and (4) for complex constituents: no language allows for contrastive ordering or conflicting vocalic features in complex onsets or codas. This rules out tautosyllabic clusters like \([ju\text{q}]\) and \([u\text{qj}]\) as well without further stipulation.

Our findings here complement and parallel those of our earlier work on aspiration and glottalization (Kehrein and Golston 2004) where we claim that:

(6) An onset, nucleus, or coda has a single unordered set of laryngeal features

That claim is based on the following observations about laryngeal contrasts in the languages of the world, parallel to (2–4) above. Given (6), we expect and find the following:

(7) No conflicting laryngeal contrasts within a margin or nucleus

\(b^p' h' hp? h?\)  \(p^h't' p^{ht}\)  are not margins in any language
\(\text{a}a \text{a}a \text{h}a? \text{?}a\text{h} \text{a}i \text{a}i\)  are not nuclei in any language

(8) No pre/post contrasts within a margin or nucleus

\(b^h\sim p^h hp\sim ph p^{ht}\sim p^{th}\)  do not contrast in margins in any language
\(\text{?}p\sim p' \text{?}p\sim p? p't\sim p't\)  do not contrast in margins in any language
\(\text{a}a\sim \text{a}a \text{h}a\sim \text{h}a \text{a}i\sim \text{a}i\)  do not contrast in nuclei in any language
\(\text{a}a\sim \text{a}a \text{h}a\sim \text{h}a \text{a}i\sim \text{a}i\)  do not contrast in nuclei in any language

(9) No segment/cluster contrasts within a margin or nucleus

\(p^h \sim ph\)  \(p^{h}\sim p\text{th}\)  do not contrast in margins in any language
\(p' \sim p?\)  \(p't\sim p't?\)  do not contrast in margins in any language
\(\text{a}a \sim \text{ah}\)  \(\text{a}i\sim \text{ahi}\)  do not contrast in nuclei in any language
\(\text{a}a \sim \text{a}a?i\)  \(\text{a}i\sim \text{a}i?i\)  do not contrast in nuclei in any language

As should be clear, limiting margins to one set of laryngeal features (6) predicts that conflicting laryngeal features will not occur even in supralaryngeally complex margins and nuclei (7), that pre- and post-aspiration are allophonic, as are pre- and
post-glottalization (8), and that orthographic distinctions like [pʰ] and [ph] play no role in phonetics or phonology (9).

The generalizations in (2–4) and (7–9) strongly suggest that laryngeal and vocalic features are not properties of speech sounds but of higher units of prosodic organization along the lines of (1) and (6). Prosodic licensing of this kind leaves the traditional notion segment pretty much gutted: if laryngeal features are licensed prosodically, then things like [pʰ, p’, l’, l] are not actually segments any longer because the aspiration and glottalization are not part of the stop or the lateral but part of a higher-order constituent; and if palatalization, velarization, and labialization are licensed prosodically, then things like [pʲ, pˠ, pʷ, pɥ, lʲ, lˠ] are not segments either, because the palatalization, velarization, and rounding are not part of the [p] or the [l] but part of what organizes the [p] and the [l]. With laryngeal and vocalic features stripped away, all that remains of the traditional segment is the physiologically necessary but not necessarily one-to-one pairing of place and manner, for example, the labial/stop pairing of [p] and the coronal/lateral pairing of [l]. Aspiration, glottalization, palatalization, velarization, and labialization are thus all features of prosodic structure above consonants; specifically, we take them to be features of the edges of syllables (onset, coda) and the edges of words.

We call the minimal pairing between place and manner a seglet and propose that a model with seglets (place/manner) rather than segments (place/manner/laryngeal/vocalic) comes closer to modeling the contrasts we find in the languages of the world. Put another way, we claim that place is licensed by manner and everything else (including manner) is licensed by prosody, so that an aspirated and palatalized [pl] is represented as a kind of bottle brush:

(10) Prosodic licensing: one set of vocalic features per margin

```
Ons
   /\    \     /
  front spread
     [pl]
```

(10) represents any onset from [phlj] to [pʰʲ] to [ʰpʰʲl], where front is the feature we assume for palatalization, spread is the feature we assume for aspiration, and [pl] is just shorthand for the features stop, labial, lateral, however those are organized within the margin in question (i.e., we remain neutral as to how the [pl] part is internally structured). Crucially, front and spread are phonologically unordered with respect to the consonantal core [pl] and the number of vocalic and laryngeal features available in a margin does not increase as the number of consonantal articulations in that margin increases: a language cannot contrast [pʰl], where the stop is palatalized but the lateral is not, with [pʰlʲ], where both the stop and the lateral
are palatalized; nor can it contrast either of those with \([p^{\text{p}}}l\)], where the lateral is palatalized but the stop is not. Phonologically, the onset is just palatalized and that is as specific as it gets; phonetic timing of palatalization with respect to the rest of the onset is done in the phonetics and has no contrastive role to play.

We may contrast this prosodic approach to a more traditional approach in which there are segments and in which each segment has its own set of vocalic features, e.g., the proposal in Clements and Hume (1995) or the theory implicit in the international phonetic alphabet (see van de Weijer 2011 for discussion). More traditional representation like this allows a vocalic feature like front to be its own segment (\(j\)) in a simple (11a) or complex onset (11c, 11d), or to be part of another segment (\(i\)), as shown in (11b):

(11) Segmental licensing: one set of vocalic features per segment (or root node)

```
<table>
<thead>
<tr>
<th>a. Onset</th>
<th>b. Onset</th>
<th>c. Onset</th>
<th>d. Onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEG</td>
<td>SEG</td>
<td>SEG</td>
<td>SEG</td>
</tr>
<tr>
<td>VOCALIC</td>
<td>VOCALIC</td>
<td>SL</td>
<td>VOCALIC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[l]</td>
<td></td>
</tr>
<tr>
<td>front</td>
<td>front</td>
<td>[l]</td>
<td>front</td>
</tr>
<tr>
<td>=([jj])</td>
<td>=([lj, jl])</td>
<td>=([lj])</td>
<td>=([jl])</td>
</tr>
</tbody>
</table>
```

The predictions of segmental (11) and prosodic licensing (10) are pretty much the same for simple glides (5a = 11a), and both accounts predict that pre- and post-palatalization do not contrast by assuming that the vocalic features within a margin (5b) or segment (11b) cannot be ordered contrastively. Both types of model also rule out a velarized palatal glide \([j^{\text{v}}}]\) and a palatalized velar glide \([u^{\text{j}}]\), by means of feature co-occurrence restrictions within a segment or syllable margin. But segmental models make markedly different predictions for glide-consonant clusters like those in (11cd). Segmental licensing of vocalic features is compatible with a three-way phonological contrast such that \([l^{\text{lat}}, ljat, jlat]\) might constitute a minimal triple in some languages, where \([lj]\), \([lj]\), and \([jl]\) are phonologically distinct onsets. The prosodic approach we propose here predicts that no language contrasts any of \([l^{\text{lat}}, ljat, jlat]\), namely, that they are either uninteresting orthographic variants allowed by IPA or are in complementary distribution or free variation in any language in which they occur. This is in line with the cross-linguistic facts we will present here.

Segmental licensing makes markedly different predictions for complex constituents of the type in (5c) as well. If every speech sound can host its own vocalic specification, we expect to find syllables like \([p^{\text{p}}}l^{\text{a}}a\) or \([p^{\text{v}}}l^{\text{j}}a\), where two consonants in the same onset have conflicting vocalic features (\(l^{\text{j}}\) requiring a front tongue position and \(v^{\text{a}}\) a back):
(12) Segmental licensing: *one set of vocalic features per segment (or root node)*

\[
\begin{align*}
\text{a. Onset} & \quad \text{b. Onset} \\
\text{SEG} & \quad \text{SEG} \\
\text{VOCALIC} & \quad \text{VOCALIC} \\
\text{SL} & \quad \text{SL} \\
[\text{front}] p & \quad [\text{back}] l \\
& = [\text{pjl}] \\
\text{SEG} & \quad \text{SEG} \\
\text{VOCALIC} & \quad \text{VOCALIC} \\
\text{SL} & \quad \text{SL} \\
[\text{front}] p & \quad [\text{front}] l \\
& = [\text{pjl}] \\
\end{align*}
\]

A traditional segmental model like (12) makes two incorrect predictions about natural language. First, it predicts that some language might allow onsets like (12a) or (12b), which seems not to be the case. Second, it predicts that some language might contrast (12a) and (12b), which is also not the case, this time *a fortiori*. Onsets like (12ab) do not occur in the languages of the world as far as we can tell, and thus constitute an overprediction on the part of segmental licensing models. We do not see how segmental models can insightfully exclude such things, so we propose to fix the problem by getting rid of segments and replacing them with place/manner pairings (seglets).

Segmental licensing also predicts contrasts like the following, where \([\text{p}^j\text{l}^a\sim\text{p}^j\text{la}\sim\text{pl}^j\text{a}]\) are a potential minimal triple:

(13) Segmental licensing: *vocalic features are licensed by individual segments*

\[
\begin{align*}
\text{a. Onset} & \quad \text{b. Onset} & \quad \text{c. Onset} \\
\text{SEG} & \quad \text{SEG} & \quad \text{SEG} \\
\text{VOCALIC} & \quad \text{VOCALIC} & \quad \text{VOCALIC} \\
\text{SL} & \quad \text{SL} & \quad \text{SL} \\
[\text{front}] p & \quad [\text{front}] l & \quad [\text{front}] p \\
& = [\text{p}^j\text{l}] & & = [\text{p}^j\text{l}] & & = [\text{p}^j\text{l}]
\end{align*}
\]

Again, our prosodic approach is not compatible with any such contrast and (13a–c) are predicted to be allophonic if they occur within a language; if no language contrasts things like (13a–c), as seems to be the case, our prosodic approach comes closer to a segmental approach in modeling what is out there.

An important point of clarification is in order here. Our prosodic approach is not about *strings* but about *constituents*: it does not rule out a contrast among \([\text{p}^j\text{l}^j], [\text{p}^j\text{l}], \) and \([\text{pl}^j]\) if the stop and lateral are in distinct words, for instance. Many languages contrast two word utterances like \([\text{kappa}\text{l}^j\text{ut}]\) vs. \([\text{kap}^j\text{l}^j\text{ut}]\) vs. \([\text{kap}^j\text{l}^j\text{ut}],\)
where the various bits of palatalization are parts of different words. And some languages contrast [ gór l], [ gór l], and [ pl l] across syllables within a word: [ ap l l a] vs. [ ap l l a]. All of this is allowed by our prosodic approach. Our claim is that no language contrasts [ gór l], [ pl l], and [ pl l] within an onset or coda, or within a word margin (prependix, appendix) in languages that have such things. As we will see below, languages with extrasyllabic consonants like the nasal in Russian mgla “mist” (ω m (σ gla)) allow palatalization of the word margin (prependix) or of the syllable margin (onset), because they are distinct constituents. This leads to near-minimal pairs like l d a “ice (gen.)” and l b e “forehead (prep.),” where the former has a palatalized word margin (ω l (σ da)) and the latter has a palatalized syllable margin (ω l (σ b e)).

We note here at the outset that our results are not meant to argue for a particular set of vocalic features. We use [front, back, round] instead of, for example, [coronal, dorsal, labial] (Clements and Hume 1995), but we expect our claims to hold either way.3 In this chapter, we focus on where the features go rather than on what the features are.

Prosodic licensing like this requires more highly structured lexical representations than simple strings of segments, of course, because laryngeal and secondary features need prosodic nodes to hang off of now that segments cannot license them. Discussion of such matters takes us beyond the scope of this chapter; we refer the interested reader to Golston and van der Hulst (1999) and Golston (2007) for proposals as to how much prosody is underlying in natural language and why. Our focus here is on possible contrasts.

The rest of the chapter is organized as follows. In section 2 we motivate our general claim that the vocalic contrasts found in languages do not increase with the consonantal complexity of the margin: complex margins like [ pl] or [ tm] show the same range of vocalic options that simple margins like [ p] or [ l] or [ t] or [ m] show, making it unlikely that each consonant in an onset hosts its own set of vocalic features. In sections 3 through 5 we substantiate the three more specific claims in (2–4) above: that languages countenance no conflicting vocalic contrasts within a syllable margin (j ̯ u̯ , * p̯ , etc.), no pre/post contrasts within a syllable margin (i p l ~ p l j, j p l ~ p j, etc.), and no segment/cluster contrasts within a margin (p l l ~ p j; pl l ~ pl j, etc.). We then consider languages that appear problematic for our proposals and generalize the analysis to the margins of prosodic words in section 6, show how prosodic licensing constrains processes of assimilation in section 7, and end with some theoretical implications of our results, including the demise of the segment and the rise of the seglet (7).

2 Complex margins are vocalically simple

Simple margins can have up to six distinct types of contrastive vocalic settings: plain (Ø, p), palatal (j, p l), velar (u̯ , p̯ ), labial (β̂ , p̂ β), labio-palatal (u̯ , p̯ ), and labiovelar (w, p̂ w) (Ladefoged and Maddieson 1996).4 These can be analyzed with three
privative features, [front], [back], and [round], along with a co-occurrence restriction against the antagonistic combination *[front, back].

(14) Vocalic features

\[
\begin{align*}
\text{VOCALIC} \\
\text{round} \\
\text{front} & \quad \text{back}
\end{align*}
\]

(15) Glides and secondary vocalics

<table>
<thead>
<tr>
<th>Round</th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>p, pʲ</td>
<td>j, pʲ</td>
<td>uŋ, pˠ</td>
</tr>
<tr>
<td>β̞, pᵢ</td>
<td>u, pɳ</td>
<td>u, pŋ</td>
</tr>
</tbody>
</table>

We start by looking at languages which allow for glide-only margins. Maddieson's (1984) database contains glides at all five places, albeit with markedly different frequencies.

(16) Approximants in Maddieson (1984)

<table>
<thead>
<tr>
<th>Places</th>
<th>Palatal</th>
<th>Labiovelar</th>
<th>Labial</th>
<th>Velar</th>
<th>Labio-palatal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of languages</td>
<td>271</td>
<td>238</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Percent of languages</td>
<td>86.1%</td>
<td>75.7%</td>
<td>1.9%</td>
<td>1.6%</td>
<td>1.3%</td>
</tr>
</tbody>
</table>

The frequencies above are roughly comparable to those of the corresponding high vowels (i, u, i/u, y) in Maddieson's database, supporting the generally accepted assumption that glides are the non-syllabic counterparts of high vowels. (Note that since vowels are always produced with some tongue body articulation, there is no syllabic sound corresponding to the labial glide [β].)

Below we give examples from languages that contrast two, three, or four glides. We are not aware of a language that contrasts five. Notice, though, that all individual contrasts seem to be attested, including contrasts of [β] and [w].
(17) Languages with two glides

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back</th>
<th>Round</th>
<th>Front, round</th>
<th>Back, round</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adzera, Bini (Edo),</td>
<td>j</td>
<td></td>
<td></td>
<td>w</td>
<td></td>
</tr>
<tr>
<td>English, Igbo, Kashmiri,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kihehe, Klamath, Korean,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kutep, Luganda, Polish,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temne, Toda</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hindi-Urdu, Karok,</td>
<td>j</td>
<td></td>
<td>β,</td>
<td>w</td>
<td></td>
</tr>
<tr>
<td>Nzima, Sámi, Telugu,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karacalar Ubykh,Yatée Zapotec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(18) Languages with three glides

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back</th>
<th>Round</th>
<th>Front, round</th>
<th>Back, round</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axininca Campa, Mazatec</td>
<td>j</td>
<td>ῑ</td>
<td>β,</td>
<td>w</td>
<td></td>
</tr>
<tr>
<td>Aranda, Cofan, Kanakuru,</td>
<td>j</td>
<td>ῑ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margi, Marshallese, Lillooet, Shuswap, Wiyot</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abkhaz, Lakkia, Twi</td>
<td>j</td>
<td>β̸</td>
<td>ῑ</td>
<td>w</td>
<td></td>
</tr>
<tr>
<td>Shona</td>
<td>j</td>
<td>v</td>
<td></td>
<td>w</td>
<td></td>
</tr>
<tr>
<td>Breton, Fante, French,</td>
<td>j</td>
<td>ῑ</td>
<td></td>
<td>w</td>
<td></td>
</tr>
<tr>
<td>Gà, Iaai, Western</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idoma, Kom, Mandarin,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tikar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(19) Language with four glides

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back</th>
<th>Round</th>
<th>Front, round</th>
<th>Back, round</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dschang</td>
<td>j</td>
<td>ῑ</td>
<td>ῑ</td>
<td>w</td>
<td></td>
</tr>
</tbody>
</table>
Margins with a single consonantal articulation occur with the same set of six secondary articulations, as we see in (20–23). Note the complete absence of languages that contrast only plain and velarized margins ($p \sim p^\text{ˠ}$):

(20) Two series of vocalically simple margins

<table>
<thead>
<tr>
<th></th>
<th>Plain</th>
<th>Front</th>
<th>Back</th>
<th>Labial</th>
<th>Front, labial</th>
<th>Back, labial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgarian, Irish, Kashmiri, Lithuanian, Nenets, Ocaina, Polish, Russian, Resigaro</td>
<td>$p^{(v)}$</td>
<td>$p^j$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(21) Three series of vocalically simple margins\textsuperscript{10}

<table>
<thead>
<tr>
<th>Plain</th>
<th>Front</th>
<th>Back</th>
<th>Labial</th>
<th>Front, labial</th>
<th>Back, labial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amuzgo, Bura, Hausa, Igbo, Kam, Lai, Lakkia, Luganda, Margi, Nambakaengo, Tera, Tsimshian, Zoque</td>
<td>p</td>
<td>p\textsuperscript{j}</td>
<td></td>
<td></td>
<td>p\textsuperscript{w}</td>
</tr>
<tr>
<td>Ubykh</td>
<td>q</td>
<td>q\textsuperscript{j}</td>
<td>q\textsuperscript{β}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scottish Gaelic, Northern Irish, Nupe</td>
<td>l</td>
<td>l\textsuperscript{j}</td>
<td>l\textsuperscript{y}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late, Nzima</td>
<td>t</td>
<td></td>
<td>c\textsuperscript{c\textsuperscript{w}}</td>
<td></td>
<td>η\textsuperscript{w}</td>
</tr>
<tr>
<td>Marshallese</td>
<td>n\textsuperscript{j}</td>
<td>n\textsuperscript{y}</td>
<td></td>
<td>n\textsuperscript{w}</td>
<td></td>
</tr>
</tbody>
</table>

(22) Four series of vocalically simple margins\textsuperscript{11}

<table>
<thead>
<tr>
<th>Plain</th>
<th>Front</th>
<th>Back</th>
<th>Labial</th>
<th>Front, labial</th>
<th>Back, labial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birom, Kutep, Twi</td>
<td>p</td>
<td>p\textsuperscript{j}</td>
<td>p\textsuperscript{β}</td>
<td>p\textsuperscript{r}</td>
<td></td>
</tr>
<tr>
<td>Mazatec</td>
<td>t</td>
<td>t\textsuperscript{j}</td>
<td>t\textsuperscript{y}</td>
<td>t\textsuperscript{β}</td>
<td></td>
</tr>
<tr>
<td>Shona</td>
<td>t</td>
<td>t\textsuperscript{f}</td>
<td>t\textsuperscript{r}</td>
<td></td>
<td>m\textsuperscript{w}</td>
</tr>
<tr>
<td>Mandarin</td>
<td>t\textsuperscript{s}</td>
<td>t\textsuperscript{c}</td>
<td></td>
<td>t\textsuperscript{c\textsuperscript{r}}</td>
<td>t\textsuperscript{s\textsuperscript{w}}</td>
</tr>
</tbody>
</table>

(23) Five series of vocalically simple margins\textsuperscript{12}

<table>
<thead>
<tr>
<th>Plain</th>
<th>Front</th>
<th>Back</th>
<th>Labial</th>
<th>Front, labial</th>
<th>Back, labial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kom</td>
<td>t</td>
<td>t\textsuperscript{c}</td>
<td>d\textsuperscript{y}</td>
<td>t\textsuperscript{β}</td>
<td>t\textsuperscript{c\textsuperscript{f}}</td>
</tr>
</tbody>
</table>

The tables above require some comment: first, not all languages have all secondary vocalic articulations with every consonant: we use “p” as a cover-symbol for consonants at various points of articulation to simplify the presentation. Second, palatalization and velarization often shift primary (consonantal) places, such that coronals [t, t\textsuperscript{s}, s], for instance,
have palatalized variants at postalveolar [t̠, tʃ, ʃ] or alveolo-palatal places [tc, .pag], but retroflexes [ʈ, ʂ, s] as their velarized counterparts (Hall 1997; Kochetov 2002, among others). Likewise, palatalized [k, x, y] can be produced as palataloalveolars [t̠, tʃ, ʃ] or as palatals [c, ç, n]; and velarized velars can show up as uvulars [q, ɣ, n]. Notice also that our treatments of Mandarin and Mazatec are not the standard ones: we assume that nucleic glides in the standard treatment are actually part of the onset, as argued for Mandarin by Duanmu (2000: 480) and for Mazatec by Golston and Kehrein (1998) (contra Pike and Pike 1947; Steriade 1994); we come back to the discussion of syllable structure below.

When we turn to consonantally complex onsets and codas like [pl] or [tm], we find that they allow the same or fewer vocalic contrasts as consonantly simple margins do: the addition of extra consonantal articulations within a margin does not open up additional secondary vocalic possibilities for that margin. This is an unexpected finding from a segmental perspective, and it strongly suggests that there is a single set of vocalic features per syllable margin (1), rather than a single set of vocalic features per segment.

Thus, Irish (Ní Chiosáin 1999), Lithuanian (Ambrazas 1997), and Russian contrast plain/velarized consonants [t] with palatalized consonants [t̠] as well as plain [st] and palatalized onset clusters [s[t]], as shown in (24). (The full picture is slightly more complicated in Lithuanian and Irish, and much more complicated in Russian. We return to these languages in later sections to give a more thorough description of how palatalization is realized in different types of complex onsets and codas.)

(24) Series of vocalized complex onsets

<table>
<thead>
<tr>
<th>Plain</th>
<th>Front</th>
<th>Labial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irish, Lithuanian, Russian</td>
<td>s[t̠]</td>
<td>s[t̠]</td>
</tr>
<tr>
<td>Kashmiri</td>
<td>mp</td>
<td>m[l]p</td>
</tr>
<tr>
<td>Kabardian</td>
<td>px tx</td>
<td>p[ɣ][ɣ] t[ɣ][ɣ]</td>
</tr>
<tr>
<td>Abadzakh</td>
<td>sk sq sx s[ɣ]</td>
<td>s[ɣ][ɣ] q[ɣ]</td>
</tr>
</tbody>
</table>

Kashmiri (Bhaskararao et al. 2009) has plain consonants [p] and palatalized consonants [p[ɣ]] as well as plain [mp] and palatalized coda clusters [m[p[ɣ]]], but nothing more detailed than that, in part because "palatalization spreads across the whole of the consonant stretch to which it is attached" (Bhaskararao et al. 2009: 14). Kabardian (Catford 1972; Kuipers 1960; Henderson 1970) and Abadzakh (Paris 1989) have plain [p] and labialized [p[ɣ]] as well as plain [px] and labialized onset clusters [p[ɣ][ɣ]]. Again, complex margins do not open up additional vocalic possibilities that simplex margins lack: we do not find the tricky four-way contrasts that segmental licensing predicts [s[ɣ]∼s[ɣ][ɣ] ∼ s[ɣ][ɣ] ∼ s[ɣ][ɣ][ɣ]], only the simple two-way contrasts that prosodic licensing predicts [s[ɣ]∼s[ɣ][ɣ]].

Limitations of space keep us from rehearsing such facts for all of the languages with secondary articulations and consonantly complex margins, but the examples
above are entirely representative of the facts generally. Aside from a few cases we
discuss below, we know of no languages in which there is any reason to think that
complex margins allow more secondary articulations than simplex margins.

3 No conflicting vocalic contrasts within a margin

Despite an extensive search we have been unable to find a single language in which
palatalization and velarization occur within the same onset or coda. This is expected
for simple margins because even standard theory posits only a single set of vocalic
features per consonant. This rules out *[jpl] *[mβ], and the like as simple margins.
More interesting is the lack of non-superscript margins like *[jpu] and *[mju],
*[jµ], and of margins with more than one consonantal articulation like *[pβl] and
*[pβµ]. To rule these out we need to restrict the vocalic possibilities of complex
margins to those of simple margins, as proposed here.

Languages do not seem to combine palatal (front) and velar (back) glides within a
single onset or coda *[jµ]. There appear to be a few counterexamples to this claim, but
they all turn out to be due to mere orthographic conventions combining “w” with “w” to
represent [µ] or some similar sound. In other words: in these cases, “w” marks a labial
but not a velar articulation, so the sound is [round, front], but not *[round, front, back].
Zoque (Wonderly 1951), for instance, has /j/ and /w/ (the latter described as “bilabial,
rounded,” p. 107) and “wj” as well, which however represents something like /βj/: “The
cluster wy is actualized as an unrounded bilabial spirant with the tongue in palatal posi-
tion” (Wonderly 1951: 107). Whatever “wy” is in Zoque, the phonetic description
makes it clear that the sound under question is palatal/front but not velar/back.

Similarly, Lakkia (Haudricourt 1967) is said to have /w/, /j/, and /jw/. Again, the lat-
ter does not represent a combination of palatal and velar articulations, which we rule
out in any margin; rather “w est une labialization, j une palatalization” (p. 169, footnote
1). Lakkia <w> is a labial glide /β/, and <jw> is just a rounded palatal /η/. Bzyb and
Southern Abkhaz (Chirikba 1996) are said to have /j, w/, and /jw/, but again <w> is a
purely labial glide /β/, and <jw> is actually /η/: “The symbol η represents a labial plus
palatal semivowel, exactly like the initial French sound of huit” (Catford 1977: 291).

Klamath presents a slightly different case (Barker 1964). The language has /w/
and /j/ as well as a word-initial “cluster” /w+j/. In this case, however, [w] is syllabic,
realized as [wu] according to Barker, and the velar and palatal articulations are het-
erosyllabic. All of this is summarized in (25).

(25) Apparent glide clusters in single margins

<table>
<thead>
<tr>
<th></th>
<th>Orthographic</th>
<th>Phonetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoque</td>
<td>wy</td>
<td>βj</td>
</tr>
<tr>
<td>Lakkia</td>
<td>jw</td>
<td>η</td>
</tr>
<tr>
<td>Abkhaz</td>
<td>jw</td>
<td>η</td>
</tr>
<tr>
<td>Klammath</td>
<td>wj</td>
<td>w,j</td>
</tr>
</tbody>
</table>
As for secondary palatalization and velarization, we are not aware of a counterexample to our claim in (2). This comes as a real surprise from the perspective of segmental licensing, for nothing in segmental licensing explains why a language with palatalized and velarized consonants and consonantally complex margins should not have, say, [pʰʲtˠ], [pʲj], [ʲptˠ], or the like. Our claim in (1) that each margin has a single unordered set of vocalic features captures this right away, along with the lack of margins like [jw], [jɰ], [ɥɰ], and their ilk.

4 No pre/post contrasts within a margin

Secondary articulations of consonants (aspiration, ejection, palatalization, velarization, etc.) are conventionally written with superscripts after the respective consonant symbol [pʰ, p’, pʲ, pˠ], presumably because these features are “often more apparent at the release than at the formation of a primary constriction” (Ladefoged and Maddieson 1996: 363). This is not to say, of course, that secondary articulations cannot start before the primary constriction is formed. In this section, we look at a few languages in which we find secondary vocalic articulations before, after, and overlapping the primary consonant articulation. In no case are these timing differences contrastive, per our claim in (3).

(26)–(28) show different timings of labial and palatal gestures of Russian [pʲ] in intervocalic, word-initial, and word-final position, respectively (from Kochetov 1999). The intervocalic case shows the palatal articulation slopping over both sides of the labial closure in a simultaneously pre- and postpalatalized stop [ʲpʲ]:

(26) Intervocalic [aʲpʲa] (Kochetov 1999:182)

```
<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lips</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TB</td>
<td></td>
<td>Palatal</td>
<td>Pharyngeal</td>
</tr>
</tbody>
</table>
```

The word-initial case (27) has audible palatalization only following the release of the stop, hence postpalatalized [pʲ]:

(27) Word-initial [pʲa] (Kochetov 1999:183)

```
<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lips</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TB</td>
<td></td>
<td>Palatal</td>
<td>Pharyngeal</td>
</tr>
</tbody>
</table>
```
Gestural overlap in word-final position (28) resembles the intervocalic case, but the amount of audible postpalatalization depends on whether and how much the labial closure is released in this position: “[t]he palatal glide at the right edge is devoiced and turned into a short component [ç], which represents an audible friction” (Kochetov 1999: 183, after Jones and Ward 1969). Thus, word-final /pʲ/ in Russian is regularly prepalatalized and postfricated [ʲpʲ]:

(28) Word-final [aʲpʲ] (Kochetov 1999:183)

The acoustic effects of such different timing options can also be seen in spectrograms from Hupa [ʲkʲʼ] and Russian [pʲ] and [ʲt].

(29) Hupa [ʲkʲʼ] (Gordon 2001: 32) (arrow marks palatal transitions into the velar closure)
As Timberlake (2004: 39) notes, prepalatalization in Russian has a clearly observable effect on the preceding vowel:

Stressed vowels, then, are affected by adjacent consonants in a consistent fashion. Before a following palatalized consonant, all vowels are fronted and/or raised, in the last third of the vowel and especially in the final transition. After a soft consonant, vowels are fronted and/or raised in the first third. Between soft consonants, vowels are fronted and raised in both transitions and, in an additive fashion, in the middle of the vowel as well.

Marshallese (Choi 1992) is another language with simple margins that are simultaneously pre- and postvocalized. The language has a vertical vowel system /i, ə, u/, three glides /j, ɰ, w/, and a simple system of basic consonants: /p, t, k, m, n, ŋ/, l, r/. Only the velars /k, ŋ/, however, may surface without secondary vocalic articulations; all other consonants are either palatalized, velarized, or labiovelarized:

<table>
<thead>
<tr>
<th></th>
<th>Bilabials</th>
<th>Coronals</th>
<th>Velars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>Palatalized</td>
<td>pʲ</td>
<td>tʲ</td>
</tr>
<tr>
<td></td>
<td>Velarized</td>
<td>pˠ</td>
<td>tˠ</td>
</tr>
<tr>
<td></td>
<td>Rounded</td>
<td>mʲ</td>
<td>nʲ</td>
</tr>
<tr>
<td>Nasals</td>
<td>Palatalized</td>
<td>mʲ</td>
<td>nʲ</td>
</tr>
<tr>
<td></td>
<td>Velarized</td>
<td>mˠ</td>
<td>nˠ</td>
</tr>
<tr>
<td></td>
<td>Rounded</td>
<td>mˠ</td>
<td>nˠ</td>
</tr>
</tbody>
</table>
Secondary features of consonants show distinct coarticulatory effects on neighboring vowels: vowels are fronted next to palatalized consonants, retracted next to velarized consonants, and retracted-and-rounded next to labiovelarized consonants. Importantly, as shown below, coarticulation is both perseverative and anticipatory in the coda, that is, while initial consonants are postpalatalized \([p^ɨ]\), postvelarized \([p^ˠ]\), and postlabiovelarized \([k^ʷ]\), final consonants by and large show the secondary articulations on both sides of the closure:

(32) Vowel qualities in C\(^{i}\)VC\(^{y}\) words (asymmetric contexts; Choi 1992: 16)

<table>
<thead>
<tr>
<th></th>
<th>Palatalized</th>
<th>Liquids</th>
<th>Velarized</th>
<th>Glides</th>
<th>Velarized</th>
<th>Rounded</th>
<th>Palatalized</th>
<th>Glides</th>
<th>Velarized</th>
<th>Rounded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(j)</td>
<td>(l^ɨ)</td>
<td>(r^ɨ)</td>
<td>(j)</td>
<td>(l^w)</td>
<td>(r^w)</td>
<td>(w)</td>
<td>(u_q)</td>
<td>(u)</td>
<td>(w)</td>
</tr>
</tbody>
</table>

We see for instance that velarized stops in the coda (left-most column in (32)) have back onglides, \([\mu, \upsilon, \omega]\), while palatalized stops have front onglides \([i, e, \varepsilon]\) (top right), and labialized stops have rounded onglides \([u, o, \omega]\) (bottom right), all of this in addition to the \([\gamma, j, w]\) at the point of release. Ladefoged and Maddieson (1996: 358–360) show a similar case of a simultaneous pre- and post-labiovelarized consonant in Pohnpeian \([\omega p^w]\). Again, the timing differences are not contrastive in any of these cases.

Some languages have pre- but no post- secondary vocalizations. Estonian (Lehiste 1965; Asu and Teras 2009) has palatalized coronals and coronal clusters, though only in postvocalic position (word-medially and finally). Palatalization is phased early with respect to both single consonants and to clusters. “Estonian has prepalatalization: palatalization occurs before rather than after the consonant and is characterized by a longer i-like transition from vowel to consonant and a quality change in the first part of a single or geminate consonant or consonant cluster” (Asu and Teras 2009: 368).

Higi (“Kamwe”, Mohrlang 1972) has labialized stops, affricates, and fricatives in onsets, all realized with prelabialization, such that the vocalic quality passes all the
way through the stop closure as it were to emerge on the other side. Labialized stops and affricates are realized as coarticulated labial-dental stops in Higi and need not concern us here; but the fricatives are realized with prelabialization of the kind we are interested in, with an onglide to the fricative:  

(33) Prelabialization in Higi (Mohrlang 1972)

<table>
<thead>
<tr>
<th>Stops</th>
<th>Africatives</th>
<th>Fricatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ptá] ‘skin’</td>
<td>[ptsi] ‘grass’</td>
<td>[wsi] ‘thing’</td>
</tr>
<tr>
<td>[b̪dí] ‘to pour’</td>
<td>[bdzi] ‘strand’</td>
<td>[wza] ‘farming’</td>
</tr>
</tbody>
</table>

Palatalization in Crow passes through the consonant from the other direction (Graczyk 2007: 13 and first author’s fieldwork) to surface as postpalatalization after a palato-alveolar consonant (ʧ, ʃ) or front vowel (i, e):

(34) Postpalatalization in Crow (Graczyk 2007)

/íkaa/ [íkʲaa] ‘see’
/ihká/ [ihkʲá] ‘egg’
/éehk/ [éehkʲ] ‘that’
/hát[ka]/ [hát[kᵃ] ‘all’
/áaf[ka]/ [áaf[kᵃ] ‘testicles’

(Note that an intervening [h] has no effect on the process and that acute accent marks high tone, not stress.)

Place and manner specifications of consonants are another source of different gestural timings in secondarily articulated consonants: with stops, nothing is audible during oral closure and thus some phase of a vocalic articulation will have to precede [ʷt] or follow [tʷ] oral closure in order to be perceived. For other consonants, however, both gestures can be perceived simultaneously, such that [nʷ, lʷ, sʷ] do not necessarily require vocalic onglides or off-glides (though they typically have them). As for consonantal place, there is a difference between [pʲ], produced with two independent articulators (lips and tongue), and [pʰ], [tʰ], or [kʰ], using the same or at least anatomically joined articulators. Palatalization of coronals [tʲ, sʲ, nʲ] and velars [kʲ, xʲ, ɲʲ] often results in a shift of the primary articulator, from alveolar to palato-alveolar [ʧ, ʃ, ɲ] or alveopalatal [ʨ, ɕ], and from velar to palatal [c, ç, ɲ] – with slight or no audible off-glides (similarly for velarization of coronal to retroflex and of velar to uvular).

Crucially for present purposes, none of these timing differences is used to form contrasts in any language we know of: a language can have postvocalized single margins [pʲ, nʷ, tˠ], or prevocalized single margins [pʰ, wⁿ, ʰt], or both simultaneously [pʰ, Ṽnʷ, Ṽt], but the pre-post issue is always allophonic, depending on the position, e.g. [pʰ¹pʰ¹pʰ¹pʰ¹], or the type of consonant involved, for example, [pʲ] vs. [ʃ]. Again, the contrasts we find in languages are compatible with a syllable margin having a single set of secondary vocalic features that may precede, overlap, or follow the primary consonant articulation.
Some languages use more than one phasing option in complex constituents, for example, they have prepalatalized and postpalatalized complex margins, or prevolarized and postvelarized complex margins. As with simple constituents, however, these timing differences are not contrastive but always a matter of phonetic variation depending on syllable position or the type of consonant involved, as far as we know. We start with the set of complex onsets in Irish:

(35) Complex onsets in Irish (from Ní Chiosáin 1999: 555ff.)

<table>
<thead>
<tr>
<th></th>
<th>Plain</th>
<th>Palatalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop+l</td>
<td>pl bl tl dl kl gl</td>
<td>p\l j b\l j t\l j d\l j k\l j g\l j</td>
</tr>
<tr>
<td>Stop+r</td>
<td>pr br tr dr kr gr</td>
<td>p\r j b\r j t\r j d\r j k\r j g\r j</td>
</tr>
<tr>
<td>Stop+N</td>
<td>tn kn gn</td>
<td>k\n j g\n j</td>
</tr>
<tr>
<td>N+C</td>
<td>mr mn –</td>
<td>–</td>
</tr>
<tr>
<td>fricative+C</td>
<td>fl fr sp sk sl sr sn sm</td>
<td>f\l j f\r j sp\l j s\l j s\l j s\l j s\l j</td>
</tr>
<tr>
<td>s+stop+liquid</td>
<td>spr spl str skr skl</td>
<td>s\l j s\r j s\l j f\l j f\r j f\l j f\r j</td>
</tr>
</tbody>
</table>

Onset clusters in Irish are generally well-behaved, either plain throughout or palatalized throughout. The four exceptions to this are [sp\l j, sm\l j, sp\r j], namely clusters of plain [s] followed by a palatalized labial stop (p or m). These clusters illustrate an important aspect of our proposal: we claim that onsets and codas are phonologically plain or palatalized (or velarized, labialized, labiopalatalized, labiovelarized), but we do not claim that every consonant in a palatalized onset cluster is palatalized. Our claim is rather that clusters with vocalic features realized early (p\l j), throughout (p\l j), or late (p\l j) do not contrast with each other. This is clearly true for Irish. The language contrasts [sp, sm, spl, spr] with [sp\l j, sm\l j, sp\r j], but it has neither *[fp, fm, fp\l j, fp\r j] nor *[fp\l j, fm\l j, fp\r j, fp\r j], let alone any of the following: *[sp\l j, sp\r j, fp\l j, fp\r j, pl\l j, pl\r j]. As with other clusters then, the contrast is between plain and palatalized onsets; but palatalization does not extend over the entire cluster if labial stops (p, m) are involved.

Kochetov (1999) shows that (word-initial) C1C2 onsets in Russian can have plain and palatalized consonants in C2 (36), but only plain consonants in C1 (37).

(36) Plain and palatalized C2 in Russian word-initial clusters (Kochetov 1999: 192–193)

p [sp\l j] ‘to sleep’ p\l j [sp\l j]atitj ‘to go crazy’
[f\p\l j]astj ‘to fall into’ [fp\l j]atero ‘five times’
t [st\l j]ado ‘herd’ [st\l j]ag ‘flag’
[ft\l j]orj ‘second’ [ft\l j]\r ‘rubbed in’
k [sk\l j]ot ‘cattle’ [tk\l j]\r ‘he/she weaves’
l [pl\l j]avat ‘to swim’ [pl\l j]aska ‘dance’
r [pr\l j]avj ‘right’ [pr\l j]amo ‘straight’
n [kn\l j]ut ‘whip’ [kn\l j]azj ‘prince’
(37) No palatalized C₁ in Russian word-initial clusters (Kochetov 1999: 193).

<table>
<thead>
<tr>
<th>C₁</th>
<th>Palatalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>*[pʲ]_avj_ 'right'</td>
</tr>
<tr>
<td>k</td>
<td>*[kʲ]_ast_ 'to put down'</td>
</tr>
<tr>
<td>t</td>
<td>*[tʲ]_ud_ 'labor'</td>
</tr>
</tbody>
</table>

We conclude from this that complex onsets in Russian are either plain (pl) or palatalized (plʲ), with late palatalization the norm, as we see in the non-exhaustive but representative set of clusters in (38).

(38) CC onsets in Russian (Kochetov 1999: 192–194; Chew 2003: 358ff.)

<table>
<thead>
<tr>
<th>C₁</th>
<th>Plain</th>
<th>Palatalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop+l</td>
<td>pl bl dl kl gl</td>
<td>plʲ blʲ dlʲ klʲ glʲ</td>
</tr>
<tr>
<td>Stop+r</td>
<td>pr br tr dr kr gr</td>
<td>prʲ brʲ trʲ drʲ krʲ grʲ</td>
</tr>
<tr>
<td>Stop+N</td>
<td>pn dn kn</td>
<td>pnʲ dnʲ knʲ tmʲ</td>
</tr>
<tr>
<td>C+fricative</td>
<td>dv, sf</td>
<td>dvʲ sfʲ</td>
</tr>
<tr>
<td>Fricative+C</td>
<td>fp ft sp st zd sk sl</td>
<td>fpʲ ftʲ spʲ stʲ zdʲ skʲ slʲ</td>
</tr>
<tr>
<td>s+stop+liquid</td>
<td>spr str</td>
<td>sprʲ strʲ</td>
</tr>
</tbody>
</table>

Irish and Russian are less different than (35) and (38) would suggest because palatalization in Russian extends phonetically to the first consonant in many of the clusters above. The factors supporting (or inhibiting) so-called “assimilation” are complex, involving phonological and sociolinguistic factors, and also some amount of free variation. According to Timberlake (2004: 61),

Whether palatalization extends over both consonants or begins in the middle of the cluster depends on the extent to which the two consonants are articulatorily linked in other respects. The more linked the two consonants, the more likely it is that palatalization will extend throughout the cluster. There is variation, and the trend is very much towards losing assimilation.

As far as “articulatory linkage” is concerned, the generalization seems to be that coronal clusters (save liquids) are usually realized with palatalization throughout (dʲnʲ, sʲtʲ, zʲdʲ, etc.), while clusters of coronal+labial (sʲpʲ, dʲvʲ, sʲfʲ) are less commonly realized with palatalization throughout (see Barry 1992; Houtzagers 2003; and Kochetov 1999, 2005 for further discussion). Again, the crucial point is that onsets in Russian use neither the position nor the extension of palatalization in distinctive ways, namely [stʲ] and [sʲtʲ] are just phonetic variants of a palatalized complex onset (stʲ).

Complex codas in Russian show an even more varied picture, though generally [r] seems to shun palatalization while other coronals attract it. But the timing of palatalization is not distinctive in these clusters either, i.e. Russian has words like ska[lʲpʲ] “scalp”, but neither *ska[lpʲ] nor *ska[lʲpʲ]; and while it has words like sko[rpʲ]
“grief”, it does not have *sko[r上司] or *sko[r上司]. The list in (39) is again representative but not exhaustive (the parenthesized superscript j in the final row refers to palatalization from assimilation).

(39) Final CC codas in Russian (Kochetov 1999: 195–197; Chew 2003: 358ff.)

<table>
<thead>
<tr>
<th>Plain</th>
<th>Palatalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>l+C</td>
<td>l</td>
</tr>
<tr>
<td>r+C</td>
<td>r</td>
</tr>
<tr>
<td>N+stop</td>
<td>mp</td>
</tr>
<tr>
<td>Stop+stop</td>
<td>pt</td>
</tr>
<tr>
<td>Fricative+C</td>
<td>sp</td>
</tr>
</tbody>
</table>

Summarizing, then, where palatalization occurs within a complex onset or coda is not generally contrastive in Russian. More generally, pre- vs. postpalatalization, pre- vs. post-velarization, and pre- vs. post-labialization are not contrastive in any language we know of. We consider a few problematic words in Russian below, but they do not seriously detract from the bigger picture.

The exact phonetic timing of consonantal and vocalic articulations within onsets and codas is beyond the scope of this chapter. Notice, however, that onsets tend to be postvocalized (pl上司), while codas are often prevocalized (a上司), something we attribute to sonority sequencing. Interfering factors, we suspect, are (i) the general extension of the palatal gesture, which seems to be longer in Irish than in Russian, for instance, as we see in Irish [p上司] vs. Russian [pl上司]; and (ii) the respective consonants involved, with [r] being the least compatible with palatalization (Hall 2000a).

5 No segment/cluster contrasts within a margin

Glides have always played two closely related roles in phonology, serving both as segments [j, uʃ, ʃ, w, ɥ] and as secondary properties of other segments in the form of palatalization, velarization, labialization, and so on: [p上司, pʲ上司, pʰ上司, pʷ上司]. With but a few exceptions examined below, the difference between a Cj cluster and a palatalized C上司 has never been claimed to be contrastive.

Palatalized or (labio-)velarized consonants are often analyzed as separate series [p上司, pʷ上司] or as clusters [pj, pw] on the basis of phonological economy or parsimony. If a language has the sounds [p, j, p上司] the latter is often analyzed as a cluster [pj] rather than a palatalized consonant [p上司], thereby simplifying the system of phonemes (e.g., Hockett 1955). Conversely, if a language seems to have palatalized consonants (C上司), one usually assumes that Cj clusters are banned. But such considerations are not without costs. A cluster analysis [pj] usually complicates the syllable structure to simplify the phoneme inventory, just as a palatalization analysis [p上司] complicates the
phoneme inventory to simplify the syllable structure. Feature economy (Clements 2001, 2003) predicts that any language with both \([p]\) and \([j]\) would prefer \([p^j]\) (which drives up the numbers of segments per feature, increasing economy) to a cluster \([pj]\) (which drives down the number of segments per feature, decreasing economy). Considerations of syllable complexity point in the same direction, since \([p^j]\) is a simple onset while \([pj]\) is complex.

But the crucial test for “Cj clusters” vs. “C^j segments,” or Cu_q vs. C^y, should be data driven, and the best way to do that is to consider contrast (the core premise behind Ladefoged and Maddieson 1996 for instance). Theories of phonology that include both Cj and C^j (or Cu_q and C^y, etc.) tacitly assume that the two contrast in some language. Except for a handful of words from Russian discussed below (section 6) we have found no language with such a contrast and therefore doubt that the issue of clusters vs. vocalized single consonants can be substantiated empirically.

6 Problematic contrasts in Russian

Before closing our discussion of existing and non-existing vocalic contrasts in onsets and codas, we would like to comment on a number of words in Russian which seem to violate our proposals in (3–4). Russian is of course well known for its complex word-initial onsets, – \([mgla]\) “haze,” \([tknut^j]\) “poke,” or \([vzgl^j]\) “look” – and since some of the problematic words involve these, we will begin by considering them first. Noting that word-internal onsets are well-behaved in terms of sonority sequencing, Yearley (1995) proposes that certain word-initial consonants fall outside of the onset proper and are licensed directly by the prosodic word:

The lack of concern for sonority sequencing on the part of these peripheral elements strongly suggests that they are in fact external to the syllable formed by immediately subsequent segmental material. Given that despite being syllable-external they are obviously still parsed (since they are audible components of the optimal form), it seems plausible that they are parsed directly by the Prosodic Word. (Yearley 1995: 546)

The sounds in \([mgla]\) “haze,” for instance, are prosodically licensed as in (40) for Yearley:

(40) Extrasyllabic consonant in \([mgla]\) ‘haze’

\[
\begin{array}{cccccc}
\text{C} & \text{C} & \text{C} & \text{V} & \text{m} & \text{g} & \text{l} & \text{a} \\
\end{array}
\]
With the [m] out of the way, the [gl] forms a normal onset to the syllable, rising in
sonority towards the vowel. This bears on our proposals in (2–4), of course, because
of the status of [m] with respect to the onset; if Yearley’s proposal is correct, (2–4)
should not apply to consonants that are licensed directly by the Pwd, as (2–4) apply
only within constituents, not across them. Although the [m] in mgla is subject to
(2–4) and the [gl] is subject to (2–4) as well, the non-constituent [mgl] is not,
more than heterosyllabic sounds would be, or sounds in adjacent words.

There are no systematic exceptions to our (3) in Russian or in any other language
we know of, but we do find in Russian a handful of exceptional lexical items that are
prima facie violations. Consider our (3), repeated here as (41).

(41) = (3) No pre/post contrasts within a syllable margin

\[ \begin{align*}
\text{j}p & \sim \text{p}^j & \text{jp} & \sim \text{pj} & \text{jp}^t \sim \text{p}^j \text{t}^j & \text{pt} \sim \text{p}^j \text{t}^j & \text{pt}^j \sim \text{p}^j \text{t}^j \\
\text{\text{y}} & \text{p} & \sim \text{p}^y & \text{up} & \sim \text{pu} & \text{yp}^t \sim \text{p}^y \text{t}^y & \text{pt} & \sim \text{p}^y \text{t}^y & \text{pt}^y \sim \text{p}^y \text{t}^y
\end{align*} \]

do not contrast in margins in any language

If the laterals in the words in (42) are part of the syllable onset, they provide coun-
terexamples to our (3). The first two parts of (3) are left intact (\(\text{jp} \sim \text{pj}\) and \(\text{jp} \sim \text{pj}\)), but
the third (\(\text{jp}^t \sim \text{p}^j \text{t}^j \sim \text{pt} \sim \text{p}^j \text{t}^j \)) is violated if the palatalization in a complex onset can
be either early (\(\text{lj} \text{d}\)) or late (\(\text{lb}^j \text{e}\)).

(42) \(\text{lj} \text{d} \text{a} \) ‘ice (gen.)’
\(\text{lb}^j \text{e} \) ‘forehead (prep.)’

But if the laterals in (42) are extrasyllabic (43), as required by sonority sequencing,
these words are no longer counterexamples to our claim: palatalization is part of the
word margin in \(\text{lj} \text{d} \text{a}\) and part of the syllable margin in \(\text{lb}^j \text{e}\). We can graph the
difference as follows:

(43) Exceptional initial clusters and extrasyllabicity

\[ \begin{align*}
\text{Psd} & \quad \text{Psd} \\
\sigma & \quad \sigma \\
\text{O} & \quad \text{O} \\
\text{N} & \quad \text{N} \\
\text{l}^j & \quad \text{l}^j \\
\text{d} & \quad \text{d} \\
\text{a} & \quad \text{a} \\
\text{l} & \quad \text{l} \\
\text{b}^j & \quad \text{b}^j \\
\text{e} & \quad \text{e}
\end{align*} \]

‘ice (gen.)’ ‘forehead (prep.)’

We do not mean of course that the segment \([l]\) hosts the palatal in “ice” while the
segment \([b]\) hosts it in “forehead,” for we have no segments and palatalization is
hosted by margins, not by individual sounds. Instead, the palatal in “ice” belongs to
the same constituent as the \([l]\) (the word margin) while the palatal in “forehead”
belongs to the same constituent as the \([b]\) (the syllable-margin or onset); that is all
that the graphs in (43) are meant to show.
This does not solve all of the problems, though, since not all affronts to (3) involve obvious sonority sequencing violations. This is the case for the words in (44), which seem to contrast where in the onset the palatalization occurs: early $[t^jm]$, late $[t^jm]$, or both $[t^jm]$.  

(44) $tm^jin$ ‘caraway’  \hspace{2em} $t^jma$ ‘darkness’  \hspace{2em} $t^jm^le$ ‘darkness (loc.sg.)’

The phonologically regular case is $[tm^jin]$ “caraway” (see (38)), with $[t^jma]$ and $[t^jm^le]$ requiring special treatment. We claim that the initial $[t^j]$ in these words, too, is extrametrical, because the cluster is separated underlyingly by yer, as can be seen in $[t^jm^no]$ “dark,” with the yer vocalized to $[i]$. Our analysis relies on Yearley again, who argues that words with stem-initial yrs have an extrasyllabic initial sound even if it does not involve a sonority sequencing violation (see Yearley 1995: 560–567 for discussion). If Yearley is correct in this, the initial $[t^j]$ in “caraway” is part of the onset and follows the regular palatalization pattern we see in Russian (see 35 and 37), while the initial yer-induced $[t^j]$ in “darkness” and “darkness (loc.sg.)” is part of the Pwd margin, as shown in (45):

(45) Regular $[tm^j]$ and exceptional yer-induced $[t^jm, t^jm^j]$ initial clusters

The exceptional palatalization of $l^jda, l^jbe, t^jma, \text{and } t^jm^le$, then, can be directly related to independently motivated claims about the licensing of the word-initial sounds: the additional possibilities for palatalization come not from the additional consonantal articulations involved, but from the additional margin: words like “caraway” have a single margin (the onset), while words like “ice,” “forehead,” and “darkness” have two margins, the onset and the extrasyllabic margin of the prosodic word.

We found one more case of an apparent timing contrast in Russian, this time at the other end of the word. As we show in the left column of (46), palatalization is regularly realized early in $[lt]$ codas, regardless of whether the final consonant is underlyingly $/t/ \text{ (a)}$ or $/d/ \text{ (b)}$. Kochetov (1999: 196) provides two exceptional forms (and we found no others), one with late palatalization $[lt^j]$ (proželt$^j$), one with palatalization throughout $[l^jlt^j]$ ($s^jel^j$).

(46) a. $vol^jt$ ‘volt’  \hspace{2em} proželt$^j$ ‘yellow tint’  
    $pul^lt$ ‘desk’  \hspace{2em} $kul^lt$ ‘cult’  
    $k^jel^lt$ ‘Celt’

b. $kobol^lt$ ‘goblin’  \hspace{2em} $s^jel^j$ ‘herring’  
    $gerol^lt$ ‘herald’  \hspace{2em} $skal^lt$ ‘skald’
Coronal\n\nCoronals generally assimilate to following palatalized coronals in Russian, but this does not happen with liquids, allowing for near-minimal pairs like those in (46). If the word-final consonants are all in the coda here, we again face violations of (3).

We assume that [tʲ] in these words, too, is extrametrical, parallel to the sonority-driven-extrametricality in words like vnutraʲ “inside,” where the palatalized [rʲ] is extrasyllabic on sonority grounds:

(47) Regular [lʲt] and exceptional final clusters

\[
\begin{array}{c}
\text{Pwd} \\
\sigma \\
\text{O} \\
\text{N} \\
\text{C} \\
\text{v} \\
\text{o} \\
1t \\
\text{vol}ḻt
\end{array}
\quad
\begin{array}{c}
\text{Pwd} \\
\sigma \\
\text{O} \\
\text{N} \\
\text{C} \\
\text{ž} \\
\text{e} \\
1 ṯ \\
(pro)žeḻḻt
\end{array}
\quad
\begin{array}{c}
\text{Pwd} \\
\sigma \\
\text{O} \\
\text{N} \\
\text{C} \\
\text{s}̱ \\
\text{e} \\
1 ṯ \\
s̱eḻḻt
\end{array}
\]

Admittedly, however, we lack independent evidence for this claim as yet. We submit, however, that words like “yellow tint” and “herring” are utterly exceptional and need to be handled in a way that violates regular Russian phonology however one treats them. They do not therefore supply strong evidence against our claim that pre/post contrasts are not allowed within a syllable- or word margin (3).

This brings us to apparent violations of our claim in (4), repeated here as (48).

(48) = (4) No segment/cluster contrasts within a syllable margin

\[
pʲ\sim p̱ \quad ptʲ\sim pṯ \quad \text{do not contrast in margins in any language}
\]

\[
pˠ\sim p̱ \quad ptˠ\sim pṯ \quad \text{do not contrast in margins in any language}
\]

Apparent counterexamples to (4) in Russian include the words in (49)

(49) pʲotr ‘Peter’ lʲot ‘ice (nom.)’
Pjot ‘drinks’ ljot ‘pours’

Ladefoged and Maddieson (1996: 364) describe the acoustic difference of the first pair as follows:

In [pʲotr] ‘Peter’ the transition away from the palatal position, indicated by a falling F2, begins immediately on consonantal release. In contrast, in [pjot] ‘drinks’ there is a short steady state before the transition begins.

We assume that something similar holds for the second pair, which we take to be something closer to [lʲot] and [lʲjot], since palatalization is generally realized with sonorants rather than before or after them. Neither of these pairs involves sonority sequencing violations, and so neither can be reanalyzed with extrasyllabic consonants: [pʲ] and [pj] are indistinguishable onsets for our approach, as are [lʲ] and [lj]
(and [lj] for that matter), since we take an onset to have just a single set of secondary vocalic features and take the difference between [ ipv] and [j] to be orthographic and not phonological or phonetic.

An unpublished study by Moldalieva (2012), however, suggests that the onsets in these words are not [p ipv]~[pj] and [ipv]~[lj], but [p ipv]~[p] and [ipv]~[l] and that the lower case [j] in both cases resides in the nucleus rather than the onset: p ipv tr “Peter” has a palatalized onset while pjot “drinks” has a plain onset and an [jo] diphthong, with the palatal in the nucleus of the syllable rather than the onset. Moldalieva asked subjects to rank pairs of words in terms of how well they rhyme, to determine the syllabic affiliation of the medial glides. The result was three groups, words that do not rhyme, words whose rhyme is “just OK,” and words whose rhyme is excellent. An example of two words that do not rhyme is given in (50); it is bad presumably because the rhyme portion of each word is different (ot~os) and no subjects said they rhymed.

(50) 0% rhyme
  l’ot~los ‘ice~elk’

This just shows that speakers knew what a rhyme was. At the other end of the scale were words that scored almost perfectly in terms of rhyme, with a pooled 93% “excellent” response (51).

(51) Excellent rhyme (93%)
  ljot~pjot ‘pours~drinks’  p’jot~bjot ‘drinks~hits’
  l’ot~m’ot ‘ice~honey’  l’ot~gn’ot ‘ice~oppression’
  ljot~bjot ‘pours~hits’  ljot~v’jot ‘pours~twists’

These words should rhyme on any internally consistent account of where [ ipv] and [j] go. Finally, there were words whose rhymes were deemed “just OK” (52).

(52) OK rhyme (50%)
  l’ot~ljot ‘ice~pours’  p’l’ot~p’l’jot ‘flight~will pour’
  jel~el ‘fir~tree~ale’  jof~of ‘hedgehog~city name’

The words in (52) should rhyme perfectly according to the usual assumption, that [ ipv] and [j] are part of the syllable margin; they should not rhyme at all according to our proposal, that [ ipv] is part of the margin while [j] is part of the nucleus. So neither model straightforwardly captures the facts.

One way we see of understanding the 50% figure involves different ways that the participants might have understood rhyming: (i) identical material in the syllable rhyme, (ii) identical material from the last sonority peak to the end of the syllable. According to our proposal here, with [ ipv] in the margin and [j] in the nucleus, speakers’ intuitions should be split: (i) would make the rhymes bad (egg, ot~jot for “ice~pours”) and (ii) would make the rhymes good (ot~ot for “ice~pours”). According to the standard proposal, with [ ipv] and [j] in the margin, speakers’
intuitions should be unanimous: (i) would make the rhymes good (ot~ot for “ice~pours”) and (ii) would as well (ot~ot for “ice~pours”), since the last sonority peak would be the mid vowel in both cases. If our reasoning here is correct, our proposal can be made compatible with the data, but the traditional proposal cannot be. More work clearly needs to be done, but we take Moldalieva’s results as promising and as better support for our proposal than for the standard approach.

If we are correct, the syllabic affiliation of [l] is the margin while that of [j] is the nucleus (53).

(53) Monophthong vs. Diphthong

Recall Ladefoged and Maddieson’s description of the acoustic difference: in “Peter” the transition away from the palatal begins immediately, but in “drinks” there is a short steady state before the transition to the vowel. (53) would account for this by having the palatal in the onset and non-moraic for “Peter” and in the nucleus and moraic for “drinks,” a plausible distinction.

Finally, we should mention an interesting near-minimal triple that Padgett (2008: 1942 footnote 2) raises (54).

(54) p^ast^l ‘metacarpus’
    pjan.stvə ‘drunkenness’
    pi.‘a.str ‘piaster’

Padgett takes this as evidence for a lexical distinction between [l], [j], and [i], which it may well be. But our proposal here offers a different possibility, in which the distinction is palatalization in the onset (p^ast^l), palatalization in a complex nucleus (pjan), and palatalization in a simple nucleus (pi).

Summing up, our proposed universals in (2–4) are well-respected by the vast majority of Russian words and by all of the regular phonological and phonetic patterns in the language. A handful of exceptions occur that contravene the regular patterns, though, and few of these are difficult to model without stretching the notion of prosodic licensing to cover extrasyllabic consonants whose extrasyllabic status must be stipulated rather than derived. The issue deserves further research but is not, we think, fatal to our proposals.
7 Assimilation

Up to this point we have looked at attested and unattested contrasts in the languages of the world. We turn now to assimilation, to how palatalization, rounding, and backness spread when morphemes are concatenated. Phonological processes like these show that vocalizations apply to margins as a whole, rather than to individual segments (7.1), and that vocalic features are independent of consonantal features (7.2). Both generalizations support the prosodic approach we advocate here.

7.1 Assimilation across syllables

As far as we have been able to find, vocalic neutralization and assimilation across syllables always apply to onsets and codas as a whole. Clear data come from Lithuanian, Irish, and Marshallese, to which we now turn.

Lithuanian (Ambrazas 1997) contrasts plain (velarized) and palatalized simple and complex onsets, the latter palatalized throughout (55).18

(55) Plain vs. palatalized margins in Lithuanian (Ambrazas 1997: 36–39)

<table>
<thead>
<tr>
<th>Plain</th>
<th>Palatalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>[k]ürti 'to create'</td>
<td>[kʲ]ürti 'to get holes'</td>
</tr>
<tr>
<td>[s]ūsti 'to grow scabby'</td>
<td>[sʲ]ūsti 'to grow angry'</td>
</tr>
<tr>
<td>[spr]ągūlas 'flail'</td>
<td>[sʲpʲrʲ]ęsti 'to decide'</td>
</tr>
</tbody>
</table>

Word-final codas are neutralized towards the plain series in their entirety (55); it is not just the final consonant that loses its palatalization:


<table>
<thead>
<tr>
<th>Palatalized</th>
<th>Plain</th>
</tr>
</thead>
<tbody>
<tr>
<td>gu[lʲsʲ]u '(I) will lie (down)'</td>
<td>gu[l] 'to lie (down)' (clipped inf.)</td>
</tr>
<tr>
<td>švi[lʲpʲtʲ]i 'to whistle'</td>
<td>švi[lpt] 'to whistle' (clipped inf.)</td>
</tr>
</tbody>
</table>

Medial clusters that arise when morphemes are brought together must agree in palatality, via regressive assimilation from morpheme-initial onsets or front vowels (57).19


<table>
<thead>
<tr>
<th>Plain</th>
<th>Palatalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>nē[l]u '(he) would carry'</td>
<td>ne[lʲtʲ]i 'to carry'</td>
</tr>
<tr>
<td>i[lʲst]a '(he) grows tired'</td>
<td>i[lʲsʲtʲ]i 'to grow tired'</td>
</tr>
</tbody>
</table>

The data in (55)–(57) show that secondary vocalic articulations affect onsets and codas as a whole in Lithuanian: onsets are plain or palatalized (55), word-final codas
lose palatality (56), and word-medial codas agree in palatality with following onsets (57). Moreover, neutralization and agreement of secondary vocalic articulations act independently of “primary” consonantal place features, which neither neutralize word-finally nor assimilate word-medially.

Palatal agreement and place assimilation in Irish\textsuperscript{20} make clear that vocalic and consonantal place features must be altogether independent because, in this language, coda nasals can assimilate to a following dorsal without a change in vocalic palatality. In (58a, b) palatalized codas /n\textsuperscript{j}i/ turn into [ŋ\textsuperscript{j}i] before plain velar stops, while in (c, d) plain codas /n/ turn into plain [ŋ] before palatalized velar stops. Velar stops thus transfer their consonantal features to a preceding nasal, but the vocalic features of nasal and stop (i.e., of the coda and following onset) remain distinct. These facts are difficult to reconcile with feature geometrical views that assume V-Place as a dependent of C-Place (Sagey 1986, among others).

\begin{enumerate}
\item Nasal place assimilation without palatalization in Irish (Ní Chiosáin 1994: 96)
\item Neutralization and assimilation under prosodic licensing of front
\end{enumerate}
With the exception of certain identical and homorganic consonants, all full consonants juxtaposed by syntactic or morphological processes are separated by excrescent vowels of reduced and obscure quality which (insofar as it is determined) can be predicted from the consonants and neighboring full vowels. (Bender 1968: 22)

(60) Epenthesis vs. vocalic assimilation in Marshallese (Bender 1968: 22, 27)

a. /tʲəɾʷbʷalʲ/  [tʲəɾʷbʷalʲ]  ‘work’
   /rʲapʲiɬpʲilʲ/  [rʲepʲiɬpʲilʲ]  ‘to roll (intr.)’
   /mʷakmʷik/  [mʷakmʷik]  ‘arrowroot’

b. /bʷoᵻtʲaũ/  [bʷoᵻtʲaũ]  ‘which boat?’
   /nʲətʲ/  [nʲətʲ]  (no example)

7.2 “Assimilation” within syllables

Consonants and glides often start out as independent entities but end up in a single onset or coda in the course of morphophonological processes. Such cases typically raise the question of whether the output of concatenation should be treated as a cluster [pʲ] or as a single consonant [pʲ]. As should be clear by now, such a question is moot from the perspective of prosodic licensing because the two are taken to be indistinguishable. In this section we discuss data that supports this view.

Vowel-glide alternations, as in French (61) and Kihehe (62) are the most familiar examples of this sort. In both languages we find high vowels becoming glides when a following vowel forces them into the onset:

(61) Vowel-glide alternations in French (Kaye and Lowenstamm 1984)

| li   | ‘tie-3.sg.’ |
| lje  | ‘tie-inf.’ |
| lu   | ‘rent-3.sg.’ |
| lwe  | ‘rent-inf.’ |
| ty   | ‘kill-3.sg.’ |
| tqe  | ‘kill-inf.’ |

(62) Vowel-glide alternations in Kihehe (Odden and Odden 1999)

| kú-hááta | ‘to be fermenting’ |
| kw-áala | ‘to open palms’ |
| li-telekwa | ‘it (cl. 5) will be cooked’ |
| lj-eheelá | ‘it (cl. 5) will breath’ |
| i-lúma | ‘it (cl. 9) will bite’ |
| j-uúsa | ‘it (cl. 9) will come’ |

“Floating” vocalic morphemes like round in Chaha or front in Harari (Leslau 1958) present a second source for consonants and vowels joined under a single syllable margin. In Chaha, a 3rd singular masculine object is marked by labialization of the last “labializable” stem consonant. As can be seen from the examples below, labialization is treated as a secondary feature (ʷ) of the respective consonants in our source.
A Prosodic Theory of Vocalic Contrasts

(63) Labialization in Chaha (McCarthy 1983: 179)

Without object      With 3rd m. sg. object
a. ɗænæg             ɗænægʷ         'hit'
   nædæf             nædæfʷ         'sting'
   nækæb             nækæbʷ         'find'

b. nækæs             nækʷæs         'bite'
   kæfæt             kæfʷæt         'open'
   bækær             bækʷær         'lack'

c. qætær             qʷætær         'kill'
   mæsær             mʷæsær         'seem'
   mækʲær            mʷækʲær        'burn'

d. sædæd             sædæd         'chase'

Zoque (Wonderly 1951) presents another case of featural affixation. In this language, 3.sg.possessive is marked by palatalizing the first consonant (64). Unlike Chaha above, every consonant in Zoque has a palatalized counterpart.

(64) Zoque featural affixation (Wonderly 1951)

[pata]  ‘mat’  [pʲa.ta]  ‘his mat’
[kama]  ‘cornfield’  [kʲa.ma]  ‘his cornfield’
[faha]  ‘belt’  [fᵃ.ha]  ‘his belt’
[sak]  ‘beans’  [ʃᵃ.k]  ‘his beans’

However, Zoque palatalization has a wider distribution, for it also occurs with affixes that consist of more than just palatalization (65).

(65) Zoque metathesis or coalescence

poj-pa                    [po.pʲa]  ‘he runs’
ʦaj-kası                  [ʦᵃ.kʲᵃ.si]  ‘on the wine’
kuj-maj                   [ku.mʲaj]  ‘a week hence’
takaj-ʔah-u               [tᵃ.kᵃ.ʔᵃ.hu]  ‘it becomes bitter’

Wonderly treats palatalization in Zoque as metathesis, such that j-pata becomes [pjata], and so on. (1951: 118, see also Hock 1985), while Hall argues for coalescence, such that j-pata becomes [pʲa.ta] (2000b: 727). From the perspective of our prosodic account, palatalization in Zoque is neither metathesis (two segments) nor coalescence (one): once palatal glides and consonants are syllabified into a single onset, they are phonologically unordered and their phonetic timing will be the same as with underived palatalization, i.e. [pʲ] rather than [ʲp].

Finally, Isthmus Mixe (Dieterman 2008) has a prefix j- that causes palatalization of the first consonant in the stem (66a), reminiscent of Zoque. Moreover, the language has several different -j suffixes (a deverbalizer, a clause-final marker, and a transitive marker) that cause palatalization of the stem-final consonant (66b). While
the language generally forbids word-initial clusters, there is one (morphologically complex) exception, showing that onset clusters are palatalized throughout (66c); and the same can be seen for coda clusters under suffixation of -j (66d):

(66) Palatalization in Isthmus Mixe (Dieterman 2008: 33, 39)

a. [pam] ‘illness’    [p’am] ‘her illness’

b. [tuːt] ‘to lay eggs’  [tuːt’] ‘egg’

c. [ʃʲniːwːj] ‘(he) knows me’  (from /j‐ʃ‐n‐iːwːj‐j/)  

d. [mja’hpa’mna’hʃʲpʲ] ‘you heal’  (from /m‐ja’h‐pa’m‐na’h‐ʃ‐p‐j/) 

Neither metathesis nor coalescence seem to be the correct view of palatalization in Isthmus Mixe. As for metathesis, j does not really switch places with consonants in clusters; as for coalescence, it is hard to see to which consonant of a cluster j will actually merge with. Rather, palatality is associated with stem-initial *onsets* (and stem-final *codas*) as a whole, just as our prosodic account predicts. Dieterman comes to a very similar conclusion (2008: 49):

Describing secondary palatalization as an autosegmental feature obviates the need for a set of palatalized consonants on the phonemic level and does not complicate the linear consonant-vowel structures of the syllable. The phonetic manifestation of the morpheme is clearly revealed by the autosegmental approach.

8 Implications: from segments to seglets

In classical phonemic theory, ordinary onsets like [ʍ] or [pʰʲ] allow for rather different phonological analyses: [ʍ] could be a voiceless labiovelar glide /w̥/, but also a labiovelarized aitch /hʷ/, or a cluster, either /hw/ or /wh/; [pʰʲ] could be a palatalized, aspirated stop /pʰʲ/, an aspirated stop followed by a glide /pʰʲ/ or a stop with a palatalized aitch /pʰʲ/, a palatalized stop with a plain aitch /pʰ/, or a triconsonantal cluster /phj/. The number of possible analyses increases with every consonant added to the onset: [pʰʲ], for instance, could be /pʰʲlʲ/, /pʰʲlʲ/ or /pʰlʲlʲ/, and so on. And so we find in the literature debates like the following on Chilcotin:

Phonetically, it is difficult, if not impossible, to determine whether this segment is a labialized glottal stop [ʔʷ] or a glottalized labiovelar sonorant [w’]. Krauss and Leer (1981: 135) raise the same question with respect to the Proto-Athapaskan-Eyak consonant to which the above Chilcotin segment corresponds. (Cook 1983: 131 footnote 5)

Feature theory does away with some of these problems. For most models at least, there is no way to distinguish things like /ʔʷ/ and /w’/. Both consist of a single root node heading a laryngeal feature [constricted] and vocalic features [round, back], with no indication as to which feature class is superior. Bottlebrush models (67) do
an admirable job in this respect, refusing to model things differently that do not contrast in natural languages.

(67) Bottlebrush margins

Linguists sometimes argue from the perspective of phonemic economy, phonological processes, historical evidence, and the like in order to advance one or the other phonological analysis: /pʰʲ/, /pʰj/, /phʲ/, /pʰh/, or /phj/, but more often than not, arguments from different areas point to different solutions.

If our proposal in (1) is correct, /pʰʲ/, /pʰj/, /phʲ/, /pʰh/, and /phj/ are all phonologically equivalent, and the issue is a non-starter. Until we find languages that contrast such things, or treat them differently in phonological processes, there is no reason to entertain them as distinct phonological entities. The fact that the contrastive secondary articulations for a simplex onset are the same as those for a complex onset strongly suggests that it is syllable margins that license laryngeal and vocalic secondary articulations rather than segments. The universal patterns in (2)–(4) argue for the same point: it is palatality in an onset that matters in phonology, not a palatal segment vs. a palatalized segment, or a cluster Cj vs. a segment Cʲ.

Not everything that can be written in IPA contrasts in natural languages. The segmental bias of the alphabet we use and the theories of phonological organization mustered to support that alphabet have driven a wedge between how we conceive of sounds and how they are actually deployed in language. Just as voicing, aspiration, and glottalization are better modeled as features of syllable margins than as features of individual segments (Kehrein and Golston 2004; Golston and Kehrein 2013), labialization, palatalization, and velarization are better modeled as features of word and syllable margins than as features of segments.

**Acknowledgments**

We thank Gulmira Moldalieva, Natalie Operstein, Charles Cairns, and Eric Raimy for their help. None of them is responsible for infelicities or inaccuracies, which are our own.

**Notes**

1 We exclude pharyngeal glides and pharyngealized consonants from the present discussion, but assume that they too are governed by onset and coda.
We take all features to be privative and we follow the earlier standard assumption of all vowels being dorsal for reasons that will become clear below (see Halle et al. 2000, for discussion).

Notice though that feature theories which assume that palatal(ization) and velar(ization) are produced with different articulators (coronal and dorsal, respectively) will have to stipulate that both cannot cooccur in a single onset or coda. Feature models subscribing to the traditional view of all vowels being dorsal can do with a more general ban on antagonistic feature specifications *[-back][+back], or *[front][back] in privative terms.

Again, we exclude pharyngeal (or radical) coarticulation to keep the scope of this paper manageable; our prediction, of course, is that pharyngeal coarticulation patterns exactly like palatal, velar, and labial coarticulation.

Adzera (Howard 2010). Kashmiri (Bhaskararao et al. 2009), Klamath (Barker 1964), Nzima (Ladefoged 1964: β alternates between [w~ɥ]), Temne (Kanu and Tucker 2010), Karacalar Ubykh (Dumézil 1965). Other data from Ladefoged (1964) and Maddieson (1984).


Dschang (Bird 1999). Ladefoged (1964) says that the Bini (Edo) contrasts bilabial [v], velar [ɣ], labial velar [w], and palatal [j] glides, though [v] and [ɣ] seem nowadays to be treated as the voiced fricatives [β] and [ɣ] in Edoid languages.

In the tables that follow, languages are included which show a contrast at at least one place of articulation. For simplicity, we show the contrasts using labials, but this should not be taken to imply that only labials show these contrasts. Coronals and velars are given when labials fail to show the maximal number of contrasts.

Irish (Ní Chiosáin 1999), Kashmiri (Bhaskararao et al. 2009), Lithuanian (Ambrazas 1997), Nootka (Stonham 1999); others from Kochetov (2008), based on Maddieson and Precoda (1990).

Bura (Ladefoged 1964), Hausa (Schuh and Yalwa 1993; three-way contrast only before [a]) Igbo (Clark 1990), Lakkia (Haudricourt 1967), Luganda (Ladefoged 1971), Margi (Hoffmann 1963), Zoque (Wonderly 1951). Ubykh (Catford 1977), Scottish Gaelic (Ladefoged et al. 1998), Northern Irish (Ní Chiosáin 1999), Nupe (Hyman 1970), Late and Nzima (Ladefoged 1964), Marshallese (Willson 2003). Others from Kochetov (2008), based on Maddieson and Precoda (1990).

Birom (Ladefoged 1964: k ~ c ~ kβ ~ cβ), Kutepe (Ladefoged 1964: ts ~ ḷ ~ ts’ ~ ts’), Twi (De Jong and Obeng 2000: only before front vowels), Mazatec (Golston and Kehrein 1998), Shona (Mudzingwa 2010), Mandarin (Duanmu 2000).

Kom (Ladefoged 1964).

"It must be stressed that although the phonetic transcription used follows Catford in indicating labialization by the letter w after the symbols for the clustered consonants, this does not imply a separate labial glide following the cluster. Labialization when it occurs is of the whole cluster, and frequently extends also to the following vowel" (Henderson 1970: 102).

Labial stops as an extreme form of postlabialization are attested with alveolars in Abkhaz and Ubykh: /tʷ, tʰʷ, dʰʷ/ = [tp, tpʰ, db], but /kʰʷ, qʰʷ, χʰʷ/ etc. = [kʰ, qʰ, χʰ] (Catford 1977).

Mutation and eclipsis increase the number of onsets in Irish significantly, but they do not change the general picture, for such clusters, too, are either plain or palatalized (see Ní Chiosáin 1999: 557ff. for examples).
Notations like [lʲp] in (38) seem to suggest some kind of medial palatalization in Russian codas, but see (30) for a more accurate phonetic description of coda Cʲ showing pre-palatalization.

Russian has final devoicing, and thus both clusters surface as [lʲt].

Before back vowels, only. Front vowels are always preceded by palatalized consonants/ clusters.

With the apparent exception of velar stops [k, g], which are “usually not palatalized, but […] ‘transparent’ for further palatalization, e.g. [ˈalʲtʃnʲɪs] ‘alder,’ [vʲɪrʲɡdʲeː] ‘(he) made one weep’ […]” (Ambrazas 1997: 37). [Transcriptions adapted to IPA. [ˈ] and [vʲ] indicate acute and circumflex accent, respectively.]

Word-internal palatal agreement in Irish is parallel to Lithuanian. Notice, though, that there is no final neutralization in Irish, e.g. sa[nʲt] ‘saint.’

See Akinlabi (1996) for more examples and formal analysis.

Chaha has labialized velars and labials, but no labialized coronals.

Archi (Kibrik 1994: 305) presents a similar case of apparent metathesis, whereby the prefix w- switches place with a following stem-initial consonant, e.g. w-sas → [sʷas]’to catch (him) inf.I sg.’

References

A Prosodic Theory of Vocalic Contrasts


5

Segmentation and Pinky Extension in ASL Fingerspelling

Jonathan Keane, Diane Brentari, and Jason Riggle

1 Introduction

At first glance, fingerspelling as a system seems easy to segment: there are a limited number of possible segments: 26, one for each letter used to write English. These segments are executed in a temporal sequence, just like written glyphs are used in a spatial sequence. However, when looking closer, fingerspelling is just like any other language stream, with many contextual dependencies and a blending of one segment into another in actual production. There are no clean boundaries that separate any two segments as the articulators, in this case the digits on the hand, move from one configuration to the next. Additionally, as will be described here, there are some examples of configurations from one segment spanning across many segments previous and following (i.e., coarticulation). This phenomenon complicates a model of segmentation: a model of segmentation that not only allows for, but predicts the types of coarticulation seen is preferable to one that cannot.

This chapter is structured as follows: section 2 shows one example of handshape variation found in fingerspelling: pinky extension coarticulation. A large corpus of fingerspelling is analyzed, and pinky extension coarticulation is found to be conditioned by surrounding segments with pinky extension. Not every letter is equally susceptible to this coarticulation, however. This will be further explored with three case studies in section 3. Finally, a model of segmentation that accounts for this coarticulation is proposed, where segments in fingerspelling are not the entire configuration of the hand, but rather, only a subpart of the hand, the active part, that has been proposed in many models of sign language phonology.
1.1 Fingerspelling

Fingerspelling, while not the main method of communication, is an important part of ASL – used anywhere from 12 to 35% of the time in ASL discourse (Padden and Gunsauls 2003). Fingerspelling is used more frequently in ASL than in other sign languages (Padden 1991). Fingerspelling is a loanword system that has a form derived from the representation of English words through a series of apogees, each of which maps to a letter in the word. Every letter used in English has a unique combination of handshape, orientation, and in a few cases movement path (Cormier et al. (2008) among others). These are used sequentially to represent an English word. Figure 5.1 shows the handshapes for ASL. The orientation of each handshape is altered in this figure for ease of second language learning. In reality, all letters are articulated with the palm facing forward, away from the signer, except for -h-, -g- (in, towards the signer), -p-, -q- (down) and the end of -j- (to the side).

Throughout this chapter only handshape is discussed. This is not to say that orientation is not important for fingerspelling (in fact the pairs -h- and -u- as well as -k- and -p- differ only in orientation). Rather, we concentrate on handshape because the coarticulatory process is specific to handshape alone; additionally, because most letters are differentiated by handshape alone. This relationship is similar to the relationship that handshape has to core lexical items in other parts of the ASL lexicon, although here there are other additional parameters: location, movement, and non-manual markers in addition to handshape and orientation. However, a sign segment will include a stable handshape (or two, if there is a handshape change in the sign), in the same way that is expected of segments in fingerspelling.

Fingerspelling is not used equally across all word categories. Fingerspelling is generally restricted to names, nouns, and to a smaller extent adjectives. These three categories make up about 77% of fingerspelled forms in data analyzed by Padden and Gunsauls (2003). In early research many situated fingerspelling as a mechanism to fill in vocabulary items that are missing in ASL. On further investigation, it has been discovered that this is not the whole story (Padden and Le Master 1985). Fingerspelling can be used for emphasis as well as when the ASL sign for a concept is at odds with the closest English word, mainly in bilingual settings. One often cited example of the first is the use of y-e-s-y-e-s and g-e-t-o-u-t. An example of the second is a teacher fingerspelling p-r-o-b-l-e-m as in a scientific problem in a science class, to clarify that what was intended here was not an interpersonal problem, but rather the setup for a scientific hypothesis. While fingerspelling is an integral

![Figure 5.1](fs-letter-asl-fingerspelling.png)
part of ASL for all speakers of ASL, it is used more frequently by more educated signers, as well as more frequently by native signers when compared with non-native signers (Padden and Gunsauls 2003).

Finally, there is already some literature on the nativization process from finger-spelled form to lexicalized sign (Brentari and Padden 2001; Cormier et al. 2008). The phonetics and phonology of fingerspelling are in many ways related to ASL in general, because it uses many of the same articulators, but there are important differences. One major difference is that because fingerspelling is comprised of rapid sequences of handshapes, it provides an excellent area to look at the effects of coarticulation on handshape. Thus it is important that we study the phonetics and phonology of fingerspelling as well as that of ASL generally. With the exception of (Wilcox 1992; Tyrone et al. 1999; Emmorey et al. 2010; Emmorey and Petrich 2011; Quinto-Pozos 2010) there is little literature on the phonetics of fingerspelling. Wilcox (1992) looks at a very small subset of words (~7) and attempts to describe the dynamics of movement in fingerspelling. Tyrone et al. (1999) looks at fingerspelling in Parkinsonian signers, and what phonetic features are compromised in Parkinsonian fingerspelling. Emmorey et al. (2010) and Emmorey and Petrich (2011) studied the effects of segmentation on the perception of fingerspelling and compared it to parsing printed text. Finally Quinto-Pozos (2010) looks at the rate of fingerspelling in fluent discourse in a variety of social settings.

2 Pinky extension coarticulation

We have found that there is, indeed, coarticulation with respect to pinky extension (compare the two images of hands fingerspelling -r- in Figures 5.2a and 5.2b). This coarticulation is conditioned by both preceding and following handshapes that include an extended pinky, although there is a clear distinction between handshapes where the pinky is extended and the other fingers are not (-i-, -j-, and -y-) and those where the pinky is extended along with other fingers (-b-, -c-, and -f-).

![Figure 5.2](image)

Figure 5.2 Apogees from (a) D-I-N-O-S-A-U-R and (b) C-H-R-I-S.
There has been a small amount of work on coarticulation in fingerspelling specifically. Jerde et al. (2003) mentions that there is coarticulation with respect to the pinky. Tyrone et al. (1999) describe some Parkinsonian signers who blend letters together and give an example of the first two fs-letters of p-i-l-l-s being blended together. Finally, Hoopes (1998) notes the existence of pinky extension coarticulation in fingerspelling but separates it from the pinky extension that he is interested in: the use of pinky extension in core lexical items as a sociolinguistic marker.

2.1 Methods

We generated a large corpus of fingerspelled words for multiple concurrent linguistic and computer-vision projects. This is the source of all of the data presented below. It was recorded with the intent to use the data in multiple ways, and thus be as flexible as possible.

2.1.1 Data collection

Three wordlists were created. The first list had 300 words: 100 names, 100 nouns, and 100 non-English words. These words were chosen to get examples of as many letters in as many different contexts as possible, and are not necessarily representative of the frequency of letter, or letter combinations in English, or even commonly fingerspelled words. The second list consisted of 300 mostly non-English words in an effort to get examples of each possible letter bigram. The third list had the 300 most common nouns in the CELEX corpus in order to get a list of words that are reasonably familiar to the signers. The data analyzed here are only from the first word list.

So far, four deaf signers have been recorded, three are native ASL users, and one is an early learner. The ages of the signers are: 65, 58, 51, and 32. Approximately six hours of video has been recorded, which includes 5,700 words (11,400 tokens) and approximately 71,250 apogees.

The data was collected across different sessions that consisted of all of the words on one wordlist. During each session the signer was presented with a word on a computer screen. They were told to fingerspell the word, and then press a green button to advance if they felt that they fingerspelled it accurately, and a red button if they had made a mistake. If the green button was pressed the word would be repeated, the signer would fingerspell it again, and then they would move on to the next word. If the red button was pressed the sequence was not advanced, and the signer repeated the word. Most sessions were collected at a normal speed, which was supposed to be fluid and conversational, the signers were told to fingerspell naturally, as if they were talking to another native signer. For a small number of sessions the signers were asked to fingerspell at a careful speed, which was supposed to be slow and deliberate. Each session lasted between 25–40 minutes and there was a self-timed break in the middle of each session for the signer to stretch and rest.
Video was recorded using at least two cameras, both at 45-degree angles from straight on. Each of these cameras recorded video that was 1920 × 1080 pixels, 60 fields per second, interlaced, and using the AVCHD format. These files were then processed using ffmpeg to deinterlace, crop, resize, and re-encode the video files so that they were compatible with the ELAN annotation software (Crasborn and Sloetjes 2008).

In order to quantify timing properties of the fingerspelled words, we needed to identify the time where the articulators matched the target for each fs-letter in the word. In other words, we needed to segment the fingerspelling stream. We will use the term handshape to refer to the canonical configuration of the articulators for each fs-letter and the term hand configuration to refer to the actual realization of handshape for a specific fs-letter in our data. We call the period of hand configuration and orientation stability for each fs-letter an apogee (i.e., where the instantaneous velocity of the articulators approached zero). This point was the period where the hand most closely resembled the canonical handshape, although in normal speed the hand configuration was often very different from the canonical handshape. For now, apogees can be thought of as the segments of fingerspelling. We will refine our definition of what constitutes a segment in section 3.

2.1.2 Timing annotation
So far we have annotated a total of three hours of video across four sessions and two different signers. This set contains 15,125 apogees, of which 7,868 are at a normal conversational speed. This is the data that was used in the pinky extension and case studies that will be discussed below.

Once the video was processed, three to four naïve human coders identified the approximate time of each apogee while watching the video at around half of the real time speed. In order to determine more precise apogee times, the apogees from each coder were averaged using an algorithm that minimized the mean absolute distance between the individual coders’ apogees. This algorithm allowed for misidentified apogees by penalizing missing or extra apogees from individual coders. Using logs from the recording session, a best guess at the fs-letter of each apogee was added using left edge forced alignment. Finally, a researcher trained in fingerspelling went through each clip and verified that this combined apogee was at the correct time, and the fs-letter associated with it matched the fs-letter being fingerspelled. A single frame was selected as the time of each apogee, even if the apogee spread over multiple frames. Most apogees are only stable for a single frame, and of those that show stability for more than one frame, it is usually only for two to three frames. Where there were multiple frames, the first frame of hand configuration and orientation stability was chosen. Where there was no perceptible hold, the frame where the hand configuration and orientation most closely matched the canonical handshape and orientation was chosen. This will introduce some noise into measurements of transition time, but for almost all apogees this noise is at most 48 milliseconds. Finally the information from these verified files was imported into a MySQL database to allow for easy manipulation and querying.
2.1.3 Hand configuration annotation

Using the timing data annotated so far, we extracted still images of every apogee. This image was associated with the corresponding apogee data in the database which not only allowed for exploratory data analysis, but was also the basis of our resulting hand configuration annotations: the still images were then used to annotate a number of different features of hand configuration. The major guiding principle in this feature annotation was to keep the task as simple and context free as possible. This has two major goals:

Simplicity – the first principle is simplicity, we wanted each annotation task to be as simple as possible. This allows the training to be simple, and the task to be incredibly quick. Rather than attempting to annotate features of hand configuration as a whole using recent annotation methods (Eccarius and Brentari 2008; Liddell and Johnson 2011b; 2011a; Johnson and Liddell 2011), we used binary decision tasks that involve looking at an image of an apogee and deciding if some feature of the hand configuration is one of two values. This makes the actual annotation very, very quick. This means that a number of annotators can be used for every apogee, which then allows us to check agreement, rate annotator accuracy, and even derive some amount of certainty or gradience about the particular phenomenon (although this gradience will not be explored or used in the current study). We defined a pinky as extended if the tip of the pinky was above a plane perpendicular to the palmar plane, at the base of the pinky finger (the MCP joint) and the proximal interphalangeal joint (PIP) was more than half extended. Note that the canonical -e- shape would not have pinky extension (Figure 5.3a), although some exhibited coarticulation (Figure 5.3b). A more nuanced definition might be needed for further work but this is sufficient to identify apogees where the pinky is not in a closed, flexed configuration. With this metric the handshapes for -b-, -f-, -i-, -j-, -y-, and sometimes -c- would have extended pinkies, and the rest of the fs-letters would not. Figure 5.3c shows a -c- without pinky extension and Figure 5.3d shows one with pinky extension. Given this definition, annotators were shown images of every apogee, and determined if the pinky was extended or not. Of course, as with all phonetic realizations, pinky extension is not actually binary. A variety of measures of the amount of extension (either for the finger overall, or individual joints) could be used, however these are all much more complicated to annotate than a simple binary decision, requiring much more annotator training and time per annotation.

![Figure 5.3](image)

Figure 5.3 Apogees from (a) E-V-E-R-G-L-A-D-E-S, (b) Z-D-R-O-Q-I-E, (c) Z-A-C-K, and (d) E-X-P-E-C-T-A-T-I-O-N.
Context free – every image was presented with as little context as possible to ensure that the annotations were as objective as possible. Annotators are likely to have a variety of biases about how canonical they expect or do not expect given hand configurations to be. In order to try and reduce the influence of annotator bias, no information was given about the apogee in the image as it was annotated. The fs-letter of the apogee was not included, nor was the word, or any features of the surrounding apogees. Although hand configurations (and orientations) that are near the handshape for a given fs-letter are easy to identify, and thus could still influence annotation decisions, hand configurations that are far from any canonical fs-letter handshape will have little to distract the annotator from the task at hand (e.g., pinky extension annotation). Additionally even if the annotator knows the hypothesis to be tested (e.g., that certain handshapes in neighboring apogees condition coarticulation), their annotation cannot be biased because they have no way of knowing what the neighboring apogees are. One possible drawback to this method is that in the case of occlusions, it is sometimes impossible to determine some hand configuration features. It is possible that in some of these cases being able to play back the contextual video would provide enough information to determine the appropriate annotation. Although this might be true for a small number of cases, the benefit of reducing annotator bias far outweighs the additional (possible) accuracy in this edge case.

2.2 Results

Looking at Table 5.1 we see that the apogees of handshapes that have pinky extension (‐b‐, ‐f‐, ‐i‐, ‐j‐, ‐y‐, and ‐c‐) by and large have it in the hand configuration as well (1,438 apogees, versus 49 apogees with no extension). Of the 49 in this set that do not have pinky extension the majority of them (36) are -c- which leaves only 13 apogees in this group. For the rest of the apogees (i.e., the handshapes that do not have pinky extension) we see a surprising 295 apogees have pinky extension, which is a bit under 5% of all apogees in this set. One source of hand configuration variation is coarticulation. In order to test if the distribution of pinky extension observed is a result of coarticulation, contextual variables around each apogee (e.g., surrounding apogee handshapes, surrounding transition times) need to be investigated.

There are numerous factors that are known or suspected to condition phonetic variation like the variation we see with respect to pinky extension. Two contextual factors are the handshape of the surrounding signs, or in this case fs-letters, as well as the transition times to and from the surrounding apogees. The hypothesis here is that surrounding fs-letters that have handshapes with pinky extension will increase the chance of an apogee’s hand configuration exhibiting pinky extension even though its handshape does not specify pinky extension. Additionally we hypothesize that if the transition between a conditioning apogee and the apogee we are interested in is faster, this will also increase the chance of pinky extension. In addition to these contextual factors there are other noncontextual factors that might affect rates of
pinky extension: the category of the word being fingerspelled (name, noun, non-English) as well as which signer is fingerspelling the word.

For a first look at the effect of the handshape of surrounding apogees we will check the two possible groups that could condition pinky extension in the hand configuration of apogees that do not have pinky extension in their handshape. The two groups of f5-letters that have pinky extension in their handshapes are -i-, -j-, and -y- as well as -b-, -c-, and -f-. For apogees with handshapes that do not have pinky extension (all f5-letters but -b-, -f-, -i-, -j-, -y-, and -c-) we see that apogees that have an -i-, -j-, or -y- on either side of them have more instances with pinky extension than those that have any other letter on either side, including -b-, -c-, and -f- (see Figure 5.4).

Using a mixed effects logistic regression with varying intercepts for the f5-letter of the apogee, as well as the specific word, we determined that the following have a significant effect on pinky extension: handshape of the apogee (of interest), handshape of the previous apogee, handshape of the following apogee, word type, and the interaction of following handshape and following transition time. Specifically, the following were correlated with an increased probability of pinky extension in the hand configuration: if the apogee of interest was a -b-, -c-, -f-, -i-, -j-, and -y- (and thus the handshape had pinky extension), if the previous or following apogee was an -i-, -j-, or -y-, if the following apogee was a -b-, -c-, or -f- (marginally), if the word type was English (as opposed to non-English), and finally if both the following apogee's handshape was -i-, -j-, -y-, -b-, -c-, or -f- and the following transition time was shorter (see Appendix for full model details).

Model predictions from the regression are visualized in Figure 5.5. Here we can see that apogees with handshapes that specify pinky extension (-i-, -j-, -y-, -b-, -c-, or -f-) almost all have pinky extension in their hand configuration as we expect (they are near ceiling). For apogees of all of the other f5-letters we can see the effect that a conditioning, surrounding apogee (f5-letter: -i-, -j-, or -y-) has on the

---

**Table 5.1** Counts for expected and observed pinky extension: where the columns are handshapes with and without pinky extension, and the rows are hand configurations with and without pinky extension. The shaded cells are those where the pinky extension in the hand configuration matches the handshape specification. Here we are using the familiar terminology observed and expected. We use the terms observed and expected, even though our hypothesis is that there is coarticulation. In other words, we are using these labels in the naïve way that we do not expect any apogee that does not (phonologically) have pinky extension in its handshape, to have it (phonetically) in its hand configuration. This set excludes 216 apogees for which there were an equal number of annotations for extended and flexed.

<table>
<thead>
<tr>
<th></th>
<th>+pinky extension</th>
<th>−pinky extension</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+pinky extension</td>
<td>1,438</td>
<td>295</td>
</tr>
<tr>
<td>−pinky extension</td>
<td>49</td>
<td>5,870</td>
</tr>
</tbody>
</table>
Figure 5.4  A plot showing the percent of apogees with hand configurations that have pinky extension, despite their handshapes not specifying pinky extension, based on surrounding handshapes. Darker colors represent a higher percentage of pinky extension.

Figure 5.5  A plot showing the effect of conditioning apogees (-i-, -j-, and -y-) on the probability of pinky extension at mean transition times for both previous and following. Dots are model predictions for an apogee with a conditioning apogee in the previous position, following position, both, or neither. The lines are two standard deviations on either side. The order of the FS-letters is based on the overall amount of pinky extension.
probability that an apogee's hand configuration will have an extended pinky. For apogees of \(f\)-letters that do not have pinky extension in their handshapes, the probability that the hand configuration is realized with an extended pinky is nearly zero if there is no \(-i\), \(-j\), or \(-y\) before or after. For some of these \(f\)-letters (in particular \(-u\), \(-g\), \(-r\), \(-l\), \(-v\), \(-k\), \(-q\), \(-o\), and \(-z\)), that probability is higher if there is an \(-i\), \(-j\), or \(-y\) apogee before or after, and increases greatly if there is an \(-i\), \(-j\), or \(-y\) both before and after.

We have found that although an \(-i\), \(-j\), or \(-y\) on either side of an apogee conditions coarticulatory pinky extension, a \(-b\), \(-c\), or \(-f\) only conditions pinky extension marginally, if at all (see Figure 5.6). The generalization is that when a pinky is extended along with other fingers (especially the ring and middle fingers), there is less coarticulated pinky extension in surrounding apogees. Although this seems like an odd distinction, it is quite natural when we look at the physiology of the hand. There are three extensors involved in finger (excluding thumb) extension: extensor indicis proprius (for the index finger), extensor digiti minimi (for the pinky finger), and extensor digitorum communis (for all of the fingers) (Ann 1993).

When extended with the other fingers there are two extensors acting on the pinky, whereas when it is extended alone there is only a single extensor. Additionally when

![Figure 5.6](image-url)
the pinky is extended and the ring finger is flexed, it must act against the juncturae tendinum which connects the pinky to the ring finger. This asymmetry results in slower, less precise pinky extension when the pinky is extended alone, compared to when the other fingers are extended with it. We suggest that it is this muscular asymmetry that accounts for the fact that -i-, -j-, and -y- condition coarticulation more than -b-, -c-, and -f-.

Although transition times do not have a large main effect, the interaction between the handshape of the following apogee and the following transition time is significant. This interaction is not surprising (quick signing or speech results in more coarticulation. See Cheek (2001) for hand configuration coarticulation in ASL), but it is surprising that there is no interaction between previous handshape and previous transition time. One possible explanation for this is that there is an asymmetry between flexion and extension of the pinky. As stated above, the pinky and ring fingers are connected to each other by the juncturae tendinum; while this ligamentous band cannot exert its own force, it connects the pinky and ring fingers, and will be stretched if the fingers are not in the same configuration (either flexed or extended) (Ann 1993). For this reason we can expect that pinky extension alone will be slower than pinky flexion alone when the ring finger is also flexed. This is because only the extension is acting against the juncturae tendinum, where as flexion would be acting in concert with it. Whereas, pinky flexion is easier when the ring finger is flexed because it relieves the tension on the juncturae tendinum, so there is no physiological force that forces the pinky to remain extended. In other words, due to the physiology of the hand we expect to see slower pinky extension, but faster pinky flexion when the ring finger is flexed. Which is confirmed in our data: we see an interaction with time for only following apogees. That is, this coarticulation is time dependent only when it is regressive, not when it is progressive.

Figure 5.7 visualizes the effect of transition time and the handshape of surrounding apogees for the fs-letter -l-. As before, the x-axis in this plot is the location of a conditioning handshape and the y-axis is the probability of pinky extension. The vertical and horizontal facets (boxes) are the z-scores of the log transformed transition times for previous and following transition times respectively. We can see that for apogees that have a conditioning handshape in either the following or both apogees, the probability of pinky extension is high at short following transition times (negative z-scores), but is much lower when the following transition time is longer (positive z-scores). Apogees that have a previous conditioning handshape do not vary much based on transition time. Finally, apogees that do not have a conditioning handshape in either apogee are near zero regardless of the transition time. The main point is that, if there is a conditioning apogee as the following apogee, the following transition time magnifies the effect of a conditioning handshape when it is short, and attenuates it when it is long (the difference between the top row and bottom row of facets, with respect to apogees with conditioning handshapes in following and both positions).

Additionally, when the word type is non-English, there is less pinky extension. This could be explained by an effect of familiarity. Both of the signers have some
familiarity with English, and thus the names and nouns chosen should not be completely unfamiliar, and some were even words that the signers fingerspell frequently in ASL discourse. The non-English words however, will not be words that the signers are familiar with, and it is expected that this will be the first time that they are fingerspelling that combination of letters. We already know that the transitions in non-English words are slightly longer (Keane 2010). It is not surprising that signers exhibited less coarticulation with non-English words beyond what is predicted by the longer transitions because of a familiarity effect. There were no significant differences between names and nouns, which also fits with data on transition times that shows little difference between these two groups. Finally, there is not a significant difference between the two signers we have data for with respect to pinky extension.

2.3 Discussion

We have seen that there does appear to be coarticulation with respect to the pinky finger: an extended pinky in a neighboring apogee will increase the probability that an apogee with no specification for pinky extension will have pinky extension in its
The previous section showed that the gestures associated with pinky extension for one apogee often spread onto the apogees that surround them. Although this is just one aspect of coarticulation, it shows that it is not possible to discreetly associate every slice of time with one, and only one apogee. Because of this, simplistic models of segmenting fingerspelling will not work: we cannot assume that every apogee’s handshape is a unit that can be separated from the context that surrounds it. Rather, a model is needed that allows for, and ideally accounts for, the coarticulation observed above. Using a phonological model of handshape that breaks the hand down into smaller units that can be controlled separately allows for such a model of fingerspelling segmentation that accounts for variability seen in some parts of the hand, but not in others.

Rather than assuming that each handshape is entirely unique, where similarities or differences between them are accidental, modern sign language phonological theories decompose each handshape into a number of features allowing for relationships to be established between handshapes based on featural similarities (Mandel 1981; Liddell and Johnson 1989; Sandler 1989; van der Hulst 1995; Brentari 1998; Eccarius 2002; Sandler and Lillo-Martin 2006). They all make use of a system of selected versus nonselected fingers to divide the articulators (digits) into groups based on what digits are active in a given handshape. The selected finger group can take on more numerous, and more complicated configurations, while the nonselected finger group will be either fully extended or fully flexed. Figure 5.8 shows how handshapes are broken into their component parts. What is important here is that the selected and nonselected fingers branch at the top, and that it is only the selected
fingers that are then composed of additional features to make contrastive joint configurations. Additionally, work on active and inactive articulators in speech has shown that inactive articulators are more susceptible to coarticulatory pressures than active articulators. The selected/nonselected or active/inactive distinction is similar to specified/(un/under)-specified distinction used by Cohn (1993). Although there might be other phenomena where these different theories would make different predictions, for the data we are looking at here they can be thought of as the same.

In addition to the overall finding that there is pinky extension coarticulation, there appears to be a tendency for some fs-letters to be resistant to pinky extension coarticulation (see Figure 5.5 reprinted here as Figure 5.9). Of the fs-letters that are not phonologically specified for pinky extension, that is, all fs-letters except for -b-, -c-, -f-, -l-, -j-, and -y-, the fs-letters with the least amount of pinky extension coarticulation, are those that do not have the pinky selected (-u-, -g-, -r-, -l-, -v-, -k-, -q-, -o-, and -z-). The fs-letters that have the pinky selected (-a-, -s-, -e-, and -o-) are all towards the lower end of pinky extension coarticulation. The fs-letters -a- and -s- stand out: both have very low rates of pinky extension. -s- does not have a single instance of an apogee with pinky extension in 217 apogees, and -a- has four apogees with pinky extension out of a total of 599. Both of these fs-letters have handshapes where all of the fingers (including the pinky) are selected and flexed. The fs-letters -e- and -o- show a bit more pinky extension than -a- and -s-, even though they ostensibly have all fingers selected as well. However, recent work (Keane et al. 2012) has shown that -e- and -o- are susceptible to a phonological process which changes which fingers are selected. This typically results in the ulnar fingers (pinky and ring) becoming nonselected, while the radial fingers (index and middle) remain selected. This trend indicates that if the pinky is flexed and selected it resists the coarticulatory pressure from surrounding extended pinky fingers. However, if the pinky is flexed and nonselected, it is more susceptible to this same coarticulatory pressure.

Definition of segments in fingerspelling. In light of the asymmetry between selected and nonselected (pinky) fingers discussed above, we propose that a segmental unit of fingerspelling is not based on the whole handshape (and orientation), but rather is the period of stability of the selected (or active) fingers. The selected fingers are
less susceptible to contextual variation and are thus more invariant than the handshape taken as a whole. The next section will go through three case studies of fingerspelled words that exhibit different aspects of the pinky extension coarticulation, as well as a case where two segments seem to overlap completely by being fused together. This is accounted for (and allowed) because there is no configuration clash between the selected fingers in the handshape of either fs-letter.

### 3.1 Case studies

Three case studies have been conducted using visual estimation of extension to examine how the articulator positions change over time, and how well that aligns with any periods of stability. For each word below, the overall extension of every finger was estimated frame by frame for the entire period of time that the signer was fingerspelling the word. An extension value of zero was defined as when the finger was fully flexed; that is when all three of the joints of the finger (the metacarpophalangeal, proximal interphalangeal, and distal interphalangeal joints) were completely flexed. An extension value of one was defined as when the finger was fully extended; that is when all three of the joints of the finger were extended completely. The thumb’s measurement of...
extension is lateral across the palm,\textsuperscript{11} with zero being on the side of the hand, negative when the thumb is crossing over the palm, and positive when it is extended away from the thumb. Although these measurements of extension are coarser than other phonetic transcription systems (i.e., that of Johnson and Liddell 2011; Liddell and Johnson 2011ab), they should be sufficient for our purposes.

Figures 5.10 and 5.11 show the extension of each finger over time for one signer, and one example of the word o-i-l. For each frame and each finger, a visual approximation of extension was made. A value of zero is the most flexed that particular finger can be, and a value of one is the most extended. Lines are given for the observed values (thick) and the expected values (thin). Additionally gray boxes extend over periods of hand configuration stability, labeled with the associated \textit{fs}-letter. For each period of handshape stability, the extension values for the selected fingers of a given

\textbf{Figure 5.10}  Still images at apogees for o-i-l.

\textbf{Figure 5.11}  Articulator trajectories for o-i-l. Gray boxes represent periods of hand configuration stability, thick lines represent observed extension (visually estimated), and the thin lines represent articulator trajectories if each apogee's hand configuration were canonical, with smooth transitions.
Segmentation and Pinky Extension in ASL Fingerspelling

fs-letter are overlaid (in darker boxes) as deviations form the dotted line at zero. We adopt the Articulatory Phonology framework, which has been used for extensively for spoken languages (Browman and Goldstein 1986, 1992; Saltzman and Kelso 1987) as well as for some sign language data (Tyrone et al. 2010). This visualization is meant to function in a way similar to the gestural scores used by Browman and Goldstein (1986); Browman and Goldstein (1992) among others. The expected values line is generated by using the extension values of both the selected and nonselected fingers from the phonological specification of a canonical version of the handshape for a given fs-letter, with spline interpolation between apogees.

Starting with the first apogee, -o-, the observed and expected extension values match. For this fs-letter, all of the fingers are selected, for the fingers, the joints are phonologically specified so that they should have about 0.5 extension, and for the thumb there should be a little bit less than zero extension. Moving on to the second apogee, the -i-, only the pinky finger is selected, which should be fully extended (ext = 1). All of the other fingers are nonselected, and should be fully flexed (ext = 0). For this apogee the observed extension for the fingers aligns with the phonological specification, the thumb, however, deviates slightly, being more extended than expected. This deviation makes the thumb more like the configuration for the fs-letter that follows it: -l-. Finally, for the last apogee, the -l-, only the index finger and the thumb are selected, both being fully extended. The rest of the fingers are nonselected, and should be completely flexed. The thumb, as well as the index, middle, and ring fingers match the expected extension values. The pinky, however, stands out: although it should be flexed, it is almost completely extended. The pinky has the same extension as the apogee before it (the -i-), an example of the coarticulation discussed in section 2. In this word, the only two deviations from expected values of extension occur with digits that are nonselected and should be extended, but are realized as more extended, being more like the configurations of surrounding apogees (the following -l- in the case of the -i- and the preceding -l- in the case of -l-).

Figures 5.12 and 5.13 show the extension over time for the word B-U-I-L-D-I-N-G. The first apogee, -b-, shows no deviation from the expected extension. The next apogee, -u-, shows no deviation for the thumb or the index or middle finger (the latter two, are selected), however the ring and pinky fingers, which are nonselected, are a little bit more extended than expected. The next apogee, the first -i-, shows a lot of deviation from expected extension values. The only digit that matches the expected extension value is the pinky, which is also the only selected finger. The ring, middle, and index fingers all are slightly more extended than expected, and the thumb is completely extended, matching the configuration of the following apogee. For the -l- apogee, the thumb and index finger are selected, and both match their expected extension values. The middle and the ring finger are slightly more extended than expected, and finally the pinky is nearly fully extended, which matches the -l- before it. In the next apogee, the -d-, the thumb as well as the index and ring finger are selected, and they all match the expected extension values. The ring and pinky fingers are nonselected; the ring finger matches the expected extension, however the pinky is much more extended than expected. Across the last two apogees the pinky is more extended than expected given the phonological specification for each
handshape, however there is a handshape with an extended pinky on either side of these two (both -1- and -s-), which is conditioning coarticulation of pinky extension. Moving on to the second -1- apogee, the pinky is selected, and matches the expected extension value. The other digits approximate their expected values, with the exception of the thumb and ring finger. Following that, the -n- apogee, has the index and middle fingers selected, both of those, along with the other digits match the expected values. There are only slight deviations of the ring and pinky fingers, both of which are not selected. Finally the last apogee, -g-, has the index finger selected,
which matches the expected extension value. Additionally all of the other digits similarly match their expected extension values. This case study shows again, that there is quite a bit of extension variation for fingers that are nonselected; especially on the pinky finger when it has apogees with pinky extension on either side. In contrast, the selected fingers of a given apogee always match the expected extension.

Moving on to a more complicated example, A-C-T-I-V-I-T-Y in Figures 5.14 and 5.15, continue to show the relationship between selected and nonselected fingers.

Figure 5.14 Still images at apogees for A-C-T-I-V-I-T-Y.
The first observed extension matches the expected extension for the first five apogees (‐a‐, ‐c‐, ‐t‐, and ‐i‐) for both the selected and nonselected fingers. After that, however, there is quite a bit of deviation: the next apogee, ‐v‐, has unexpected pinky extension, as well as some articulatory undershoot for the two selected fingers (the index and the middle finger). After that the next period of stability is actually two apogees (‐i‐ and ‐t‐) fused together to form ‐it‐. The selected fingers for these two fs‐letters do not clash: for the ‐i‐ the only selected finger is the pinky, whereas for the ‐t‐ only the index finger is selected. The two sets of selected fingers are separate, and thus do not conflict. The observed extension for the index and pinky fingers reach the extension targets for ‐i‐ and ‐t‐ at the same time, and thus the two apogees occupy the same period of time. In Figure 5.15, a period of stability has been inserted halfway between the ‐v‐ and ‐it‐ to show what the articulators are expected to do if the fusion did not occur. The last apogee, ‐y‐ matches the expected configuration. This case study shows two things: First, during the period of time between the two ‐i‐ apogees (including the fused ‐it‐ apogee), the pinky does not ever completely flex, but rather stays at least partially extended as a result of coarticulation, and the fact that it is not selected in any of the intervening apogees. Second, in some extraordinary cases apogees that do not have conflicting selected fingers can be fused temporally, where the articulators reach their phonologically specified targets at the same time.

Although rare, the apogee fusion seen here is not a solitary example. There are also examples of ‐ti‐, ‐ni‐, and ‐oi‐; the last one is even documented as one strategy that is used in rapid lexicalization (Brentari 1998). Two out of three of these share the property that the selected fingers of the two fs‐letters are distinct, and thus there is no conflict. The ‐oi‐, however seems to present a problem because a canonical ‐o‐ should have all fingers selected. There is some work (Keane et al. 2012) that shows that there are instances of ‐o‐ where the ulnar digits (typically the pinky
and ring fingers) are completely flexed rather than having the same configuration as the radial digits (typically the index and the middle fingers). This happens in approximately 25% of -o-s in this corpus. The analysis of these variants are that these handshapes have different selected fingers, the canonical forms, that is, only the index and middle fingers are selected, while the pinky and ring fingers are non-selected. Additionally the one example of -oi- shows increased flexion of the ring finger, just like with the -d- in the B-U-I-L-D-I-N-G case study, suggesting that this case of -oi- fusion might involve an -o- variant that does not have the pinky finger selected. More work, and more data, are needed to fully understand and model how these two different types of variation interact.

Given a model of segmentation that looks at handshape as a whole, these fusions would have to represent examples of new kinds of segments in the inventory of fs-letters. However, if only the selected fingers are used as a basis for segmentation, these fused apogees can still be analyzed as two apogees, that just happen to occupy the same time. Why this fusion occurs is outside of the scope of this work, however many (e.g., Wilcox 1992; Emmorey and Petrich 2011) have noted that fingerspelling often has a rhythmic pattern. We have observed what appear to be consistent rhythmic patterns in our corpus and we speculate that the fusion process might be a way to maintain the rhythm when two apogees are too close together, and do not have conflicting selected fingers. More data and analysis is required to understand this possibility fully.

### 4 Conclusions

Using data from the phonetic configuration of the pinky finger, we can explore the question of what constitutes a segment in ASL fingerspelling. The pinky shows clear evidence of phonetic variation as a result of coarticulatory pressures. We have observed that there are situations where there is a higher probability of having a phonetically extended pinky finger in handshapes that are not phonologically specified for an extended pinky. The main contextual factor is that there is more pinky extension when there is a surrounding handshape where the pinky is selected and extended, compared to when there is a surrounding handshape where the pinky is not extended. This coarticulation does not occur across the board: there is temporal variation; regressive coarticulation is time-sensitive, whereas progressive coarticulation does not seem to be. Familiar words exhibit more coarticulation. Finally, not all fs-letters exhibit pinky extension coarticulation at the same rates. A trend is observed that when the pinky finger is selected and flexed, it resists pinky extension coarticulation much more than when it is nonselected and flexed.

Because of this coarticulation, defining fingerspelling segments in discrete temporal terms is not possible; that is, the articulatory gestures associated with one apogee sometimes stretch across multiple apogees. Further, as a result of the coarticulation described here, not every articulator reaches the phonologically specified canonical target for a given handshape. Which articulators reach canonical
configuration depends on their phonological status: selected fingers typically attain their phonological specification, whereas nonselected fingers show more variation, and do not always reach their phonological target. The selected/nonselected distinction was made in terms of articulator activity or inactivity as used in Articulatory Phonology, which for this data, is broadly compatible with underspecification as it has been proposed by others. Three case studies were conducted that show how this segmentation can be implemented using ideas from work on Articulatory Phonology in spoken and signed languages. Additionally, this model of segmentation can accommodate a process of apogee (or segment) fusion that is observed in fingerspelling. During this process two apogeess are executed at the same time. This is possible because the handshapes observed here have two distinct sets of selected fingers. This being the case, there is no conflict in reaching the articulatory target for the selected fingers; in order to maintain the overall rhythmic timing the apogeess for each of the two individual fs-letters in the word are collapsed into a single period of stability. The model of fingerspelling segmentation proposed here accounts for this process of apogee fusion, as well as the effects of coarticulation being limited in selected fingers because the core of the segment is restricted to the selected fingers: which is the element of handshape that carries the most contrast and information about the segment identity. Other parts of handshape are allowed to, and do, vary to ease the articulatory effort needed to produce fingerspelling, or maintain its overall rhythmic structure.

Appendix: regression model of pinky extension coarticulation

The model was fit with pinky extension as the outcome variable, as well as all and only the predictors listed in Table 5.2. We included varying intercepts for the fs-letter of the apogee because we expect that there will be variation among different fs-letters with respect to the amount of pinky extension. We included varying intercepts for words because we expect that there is variation among words with respect to the amount of pinky extension. The model was fit on only word-internal apogeess, since the first and last apogee lack a previous and a following apogee respectively.

Acknowledgments

This project simply could not have happened without the contributions of many people. This project could not be possible without the help of the signers who fingerspelled for many hours for us. We also must thank all of the collaborators that helped in this project: Karen Livescu, Greg Shakhnarovich, Susan Rizzo, Erin Dahlgreen, Katie Henry, and Katherine Mock. Finally, we must thank people for their many helpful comments at the CUNY Phonology Forum Conference on the Segment as well as the meeting of the North East Linguistic Society 42. Although all of these people helped immensely, any mistakes or omissions are entirely our own.
Table 5.2  Mixed effects logistic regression coefficient estimates and standard errors.

<table>
<thead>
<tr>
<th>Coefficient (standard error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept −7.53 (0.82)***</td>
</tr>
<tr>
<td>apogee of interest: ‐B‐, ‐C‐, ‐F‐, ‐I‐, ‐J‐, or ‐Y‐ 16.29 (2.00)***</td>
</tr>
<tr>
<td>previous ‐B‐, ‐C‐, or ‐F‐ 0.70 (0.62)</td>
</tr>
<tr>
<td>previous ‐I‐, ‐J‐, or ‐Y‐ 4.81 (0.45)***</td>
</tr>
<tr>
<td>previous transition time (zscore of log(time)) 0.21 (0.20)</td>
</tr>
<tr>
<td>following ‐B‐, ‐C‐, or ‐F‐ 1.71 (0.82)*</td>
</tr>
<tr>
<td>following ‐I‐, ‐J‐, or ‐Y‐ 3.77 (0.42)***</td>
</tr>
<tr>
<td>following transition time (zscore of log(time)) −0.16 (0.20)</td>
</tr>
<tr>
<td>word type: foreign −1.14 (0.40)**</td>
</tr>
<tr>
<td>word type: name −1.00 (0.37)**</td>
</tr>
<tr>
<td>signer: s1 −0.13 (0.30)</td>
</tr>
<tr>
<td>previous ‐B‐, ‐C‐, or ‐F‐ × previous transition time −0.61 (0.55)</td>
</tr>
<tr>
<td>previous ‐I‐, ‐J‐, or ‐Y‐ × previous transition time −0.01 (0.27)</td>
</tr>
<tr>
<td>following ‐B‐, ‐C‐, or ‐F‐ × following transition time −3.12 (0.65)***</td>
</tr>
<tr>
<td>following ‐I‐, ‐J‐, or ‐Y‐ × following transition time −2.82 (0.36)***</td>
</tr>
</tbody>
</table>

AIC 742.60
BIC 852.32
Log likelihood −354.30
Deviance 708.60
Num. obs. 4695
Num. groups: word 300
Num. groups: verLetter 26
Variance: word.(Intercept) 1.26
Variance: verLetter.(Intercept) 7.77
Variance: residual 1.00

*** p < 0.001 ** p < 0.01 * p < 0.05.

Notes

1 Which will be defined in more detail later in section 2.1. For now, they can be assumed to be synonymous with segments.

2 Traditionally movement is said to only be used for the letters -j- and -z- as well as to indicate some instances of letter doubling. Although in fluent fingerspelling many letters have movement of some type.

3 This figure was generated using a freely available font created by David Rakowski. This figure is licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License and as such can be reproduced freely, so long as it is attributed appropriately. Contact jonkeane@uchicago.edu for an original file.

4 I am choosing to adopt the typographic conventions of Brentari and Padden (2001). Fingerspelled forms are written in smallcaps (an adaptation from Cormier et al. (2008)), with hyphens: A-T-L-A-N-T-I-C and ASL native signs are written in only smallcaps: GROUP. Single finger spelled letters will be flanked by hyphens on either side (e.g., -T-).
These are also called foreign, although that is not entirely accurate, since all fingerspelled words are, in some sense, not part of the native ASL lexicon. These words were selected specifically for sequences that are not generally found in English.

The instructions, given in ASL were to: “proceed at normal speed and in your natural way of fingerspelling.”

Again, in ASL “be very clear, and include the normal kind of transitional movements between letters.” The signers were also specifically asked not to punch the letters with forward movements, as is often done for emphatic fingerspelling.

Differentiating between handshape and hand configuration follows others (Whitworth 2011), although it uses the term hand configuration in a way that is quite different from how it is used in the Hand-Tier model (Sandler 1989).

Cheek 2001 for environment; Mauk 2003 for speed and environment; Lucas et al., 2002 for grammatical category.

Where 0 represents the mean value, −1 represents a transition that is one standard deviation shorter than the mean, and +1 represents one standard deviation longer than the mean.

This movement is, of course, not physiologically extension for the thumb (rather, it is a combination of abduction and opposition). We include it here with extension for the other digits because it is the most visually salient and distinctive configuration of the thumb.

For the thumb, the extension should be slightly less than zero because it is crossing over the palm.

What fingers are selected for the fS-letter -d- is not actually a settled matter. In some models the thumb as well as the middle, ring, and pinky fingers are selected, the index finger is either nonelected and extended, or secondary-selected. However, Keane et al. (2012) has shown that -d- is frequently realized as what is referred to as baby-d-, that is with the pinky and middle fingers completely flexed, the middle finger and the thumb forming a loop, and the index finger fully extended. The apogee here, shows this pattern with flexion in the ring finger, although the pinky is extended because of coarticulation from -i- apogees around it. With that configuration the middle finger and thumb would be selected, and the index finger secondary-selected, while the ring and pinky fingers are nonselected.

A reviewer pointed out that this phenomenon may be similar to that of vowel coalescence in Sanskrit (e.g., /i/ + /a/ > /eː/). Although in the fingerspelling case the temporal properties of the fused segment seem to match a single segment more than a double segment.

References


1 Introduction

This chapter provides a brief overview of the balance that theorists take between assuming the existence of categorical units of analysis, such as segments, and the fact that there is indisputably a large number of gradient phenomena in languages. In particular, I will be demonstrating that it is feasible to develop a model of phonological relationships that is itself gradient and that builds on the insights that have been made about gradient phonological phenomena, while maintaining the assumption of categorical units such as segments. While I do not provide specific evidence that one needs to maintain the existence of segments in order to understand phonological phenomena, I aim to show that assuming them does not preclude one from building probabilistic models. Thus, we can begin to understand complex, gradient phenomena by building models that rely on standard assumptions of phonological units, and eventually (if the evidence warrants), move away from these assumptions into a more completely gradient realm (e.g., with categorical units emerging only as artifacts from gradient data).

2 Gradience and categoricality

2.1 Gradience vs. categoricality in sound systems

As discussed at length in Ernestus (2011), the difference between gradience and categoricality is often taken to be the mainstay of the difference between phonetics and phonology, respectively (though cf. Flemming 2001). That is, phonetics is often
defined as the realm of the physical (articulatory, acoustic, auditory) aspects of lan-
guage (i.e., its implementation) and is therefore necessarily gradient in nature: artic-
ulatory gestures, acoustic waveforms, and auditory processing are non-discrete and
continuous. Phonology, on the other hand, is taken to be the realm of the abstract,
competence-related aspect of sound patterns in language and is therefore more
properly a part of “grammar.” Because of this abstraction and the focus on exam-
ining patterns in sound systems, the essential nature of phonology is generally
assumed to be categorical: there are discrete units that can be combined in various
discrete ways and acted upon by rules or constraints as part of categorical processes.

Examples of both gradience and categoricality abound in the literature, and many
of them are discussed in this volume. One of the main arguments raised against the
possibility of segments being a basic unit of phonology is the fact that it is often
extremely difficult to actually identify anything “segment”-like due to the high
degree of gradient variability in speech. Variation across languages, speakers, and
utterances means that segments have no invariant identifying characteristics, and it
has certainly been demonstrated often enough through common experience that
segmenting recordings in to their component parts is often extremely difficult, with
one segment “overlapping” or “bleeding into” another (see, e.g., Hagiwara 2009;
Beckman and Edwards 2010). At the same time, categoricality is still standardly
assumed in phonological analysis, such that phonological rules (or constraints) are
often written at the level of the segment. For example, in the December 2012 issue of
*Phonology*, four of the five full-length articles propose some kind of analysis at the
level of the segment (the fifth is about metrical stress and makes reference only to
the abstract levels of the syllable and mora). In addition to this abstract analytical
approach (which arguably does not necessarily represent a belief on the part of the
authors that segments are the right unit of analysis, but could simply be a convenient
shorthand), there is psycholinguistic evidence that units the size of segments are
psychologically real in some sense. The phenomenon of categorical perception (e.g.,
Liberman et al. 1957; Studdert-Kennedy et al. 1970; papers in Harnad 1987), for
example, shows that at least some continua of phonetic variables are perceived with
sharp boundaries at particular points along the continuum, though such results
have been found to be clearer for some phenomena, such as VOT (e.g., Abramson
and Lisker 1970), than for others, such as vowel height (e.g., Fry et al. 1962; see also
discussion in Repp 1984). The concept of normalization in speech perception, too,
is predicated on the idea that listeners pull out some kind of abstract category despite
the high degree of variability in the speech signal (e.g., Johnson 2005). While some
would argue that the apparent ability to perceive segment-sized chunks comes from
familiarity with letter-based writing systems and thus is not a true characteristic of
linguistic competence (e.g., Silverman 2006; see also discussion in Burton and Noble
2004), the very fact that letter-based writing systems exist indicates that there is
some psychological awareness of abstract, segment-sized units of speech.

Ernestus (2011) gives several examples of phonological processes of which some
reported instances appear to be categorical and others appear gradient. For Korean
place assimilation, she argues (on the basis of Kochetov and Poupil`er’s 2008
articulatory study) that the process is categorical, in that a form like /paþpða/ “rather than the field” is realized as [paþpða], with a categorical replacement of /t/ with [p] – “this assimilation results in the categorical absence of the gestures for the original articulation place of the assimilated consonant” (Ernestus 2011: 2117). While of course this example is consistent with the categorical unit of analysis being a feature, a gesture, or a segment, she also gives the example of /n/ being realized as [k] in Italian (Farnetani and Busà 1994), which involves at least a combination of gestures or features, if not a segment.1 On the other hand, Ernestus points out that not all examples of place assimilation involve such categorical changes (regardless of the way they may be represented phonologically). Ellis and Hardcastle (2002), for example, show that for some English speakers the [ŋ] in an assimilated version of “ban cuts” as [bæŋ kɔts] is not identical articulatorily to the [ŋ] in “bang comes” [bæŋ kɔmz] based on an EPG analysis.2 Processes of deletion and neutralization are also discussed by Ernestus (2011) and again shown to vary between being categorical and gradient (a classic example being that of Browman and Goldstein (1990), who show that while there may be no [t] acoustically in some productions of the phrase perfect memory, there is often an articulatory gesture toward an alveolar stop, visible by X-ray).

Further support of the utility of gradience in understanding phenomena relating to linguistic sound systems comes from the fact that language users perceive and store much fine-grained phonetic detail about particular talkers and utterances, rather than just processing the segments or the words (see, e.g., Goldinger 1996; Fisher et al. 2001). Foulkes and Docherty (2006: 411) claim that within sociolinguistics, “In the majority of cases, sociophonetic variation is gradient rather than categorical. That is, variation may be observed such that a given form is used statistically more by one social group than another, or more in one speech style than another.”

At the same time, it is also clearly the case that the extent to which variability and phonetic overlap impede the categorical identification of segments in a speech stream is highly dependent on which segments are being examined and the context they are being examined in. A [t], for example, may be quite segmentable when surrounded by vowels, but almost impossible word-initially. An [r] may color an adjacent vowel inseparably. Should the fact that there are some instances in which the “beads on a string” analogy for speech fails mean that it should never be used? In practice, it is still clearly the case that this analogy is alive and well.

In particular, models of phonology which in many ways seem to embrace gradience often do so while maintaining an assumption of abstract, categorical units. Pierrehumbert (2003: 177) makes this link quite explicit, in fact deeming it necessary: “[The] conception of phonology as a formal grammar (with abstract variables) is often assumed to stand in opposition to the idea that phonology involves statistical knowledge. However, this opposition is a spurious one, because probability theory requires us to assign probability distributions to variables. Without any variables, there would be no way for a statistical learning model to tabulate any statistics about anything. Once we have variables, they can be as abstract as we like; in principle, we
can assign probability distributions to any variable of any nature that we may care to define in response to scientific findings.”

Models of so-called frequency effects in phonology, for example, are highly dependent on there being countable categories. Ernestus (2011: 2129) gives several examples: “[P]articipants are better at repeating nonce words made up of high-frequency rather than low-frequency phoneme sequences (Vitevitch et al. 1997) and at transcribing such words orthographically (Hay et al. 2004). Participants tend to interpret ambiguous fricatives as the most probable ones given the preceding and following segments (Pitt and McQueen 1998).” Such conclusions require us to accept categorical and perhaps segment-sized units to count in order to determine whether a sequence is high or low frequency and to calculate the probability of particular segments in context. Similarly, the kinds of cases that Foulkes and Docherty (2006) refer to above, in which “a given form is used statistically more by one social group than another” does not at all preclude the possibility that the form in question is some abstract categorical unit, such as a segment. This kind of mixture of gradience and categoricality within a phonological model is one to which we shall return in section 3.

2.2 Approaches to dealing with both types of phenomena

The issue of dealing with both gradience and categoricality has been addressed by a number of researchers, and there are several different kinds of approaches that have been taken to dealing with it. Erker (2012), for example, simply accepts that both kinds of phenomena are prevalent, and models each kind separately. In his analysis of coda /s/-deletion in the Spanish spoken by New York City speakers versus Latin American speakers, Erker includes both a discrete measure of deletion (either a token contains or does not contain [s]) and a continuous measure (a combination of the number of milliseconds of [s] duration with the frequency value in Hertz of the centre of gravity of whatever fricative is present). His findings reveal that while the discrete measure shows differences between “Mainland” and “Caribbean” styles of Spanish, with the Mainlanders showing more stability over generations, the continuous measures illustrate that both Mainland- and Caribbean-Spanish speakers show similar effects of contact with English in NYC and that neither is entirely stable.

Other approaches to dealing with the dual nature of phonological phenomena have tended to take a less separatist view. The Variable Rules approach (e.g., Labov 1969; Cedergren and Sankoff 1974), for instance, adds in probabilistic components directly to the application of rules. A rule that involves the categorical changing of one segment to another might be expected to apply only 60% of the time, for example. Similarly, Stochastic Optimality Theory (e.g., Boersma 1998; Boersma and Hayes 2001) involves weighting Optimality-Theoretic constraints with the probability of their being ranked in a particular way; for example, constraint 1 may outrank constraint 2 90% of the time, but the reverse may be true the other 10% of the time. Again, the units referenced by such constraints may themselves remain categorical

Another tack is that taken in Articulatory Phonology (e.g., Browman and Goldstein 1986, 1989, 1992; Gafos and Benus 2006; Pouplier 2011). In this approach, the traditional categorical units are redefined not as abstract elements that exist of their own accord, like segments, but rather as gestures. As Pouplier (2011) points out, a gesture may happen to correspond in size to a feature or a segment or some other previously defined phonological unit, but they have no inherent size. Furthermore, gestures are themselves abstract and categorical in the sense that they may be present or absent, but allow for gradience in terms of their precise implementation and the relative timing of multiple gestures.

Other approaches involve allowing phonological categories to emerge from gradient data, through the use of exemplar models, for example (e.g., Goldinger 1997; Johnson 1997; Maye 2000; or “enhanced” exemplar models, e.g., Wedel 2006; Pierrehumbert 2006; McMurray et al. 2009; Goldsmith and Xanthos 2009; Dillon et al. 2010). Pierrehumbert (2006: 519) sums up this approach, saying “Phonological categories, as represented in the mind, are viewed as clusters of similar experiences,” which themselves are of course continuously gradient.

What all of these approaches demonstrate is that it is entirely possible to build probabilistic models of phonology even while assuming the existence of categorical segments. Furthermore, it is not necessary to assume that these categories are primitives of some sort; they could themselves be emergent. In the next section, I turn to an example of such a model that is quite flexible in terms of the (categorical) units assumed but which evaluates phonological relationships in a gradient fashion.

### 3 A probabilistic approach to phonological relationships

#### 3.1 The problem

A “phonological relationship” is the relation between any two sounds in a language, typically classified as either contrast or allophony. Sounds that can be demonstrated to be contrastive in a language are typically assumed to be phonemic, and serve to form the phoneme inventory of the language; such inventories are generally taken to be the fundamental building blocks of a language’s phonological system. For the most part, such inventories are represented as sets of segments, but one can also refer to contrastive features or contrastive suprasegmental properties; the notion of contrast is not inherently tied to the notion of the segment. Still, phonological relationships are most often defined over segments and are certainly typically defined over some sort of abstract categorical phonological unit. The metric presented below will assume segments for the sake of generality, but any sort of phonological units that are of some countable size would be possible input to the model.

The problem with the standard ways of determining phonological relationships is that they assume a binary division between contrast and allophony that does not in
fact seem to capture all of the possible ways that two sounds (features, segments, etc.) can relate to each other. Typically, two sounds are assumed to be contrastive if there is at least one instance in which they are unpredictable from the surrounding phonological context (e.g., [d] and [ð] are contrastive in English because of minimal pairs such as dough vs. though, in which both sounds occur in the context [ʌʊʊ]). Two sounds are assumed to be allophonic if it is possible in every context to predict which of the two sounds will occur (e.g., [d] and [ɾ] are allophonic in English because [ɾ] occurs always and only in words like rider, in the onset position of an unstressed syllable immediately following a stressed syllable, while [d] occurs elsewhere, in words like dough, adept, had, etc.). As even a cursory glance through the descriptive literature in phonology shows, however, there is an abundance of cases in which it is not entirely clear whether two sounds should be considered contrastive or allophonic, giving rise to a plethora of terms for apparently “intermediate” relationships, including: semi-phonemic, quasi-phonemic, quasi-contrastive, partially contrastive, marginally contrastive, marginally phonemic, semi-allophonic, and quasi-allophonic (see Hall 2013a for a more complete list and references).

There are many reasons why a relationship may not be clearly contrastive or allophonic, as Hall (2013a) catalogues. The particular type of intermediacy that will be focused on here is intermediacy of predictability of distribution. The classic assumption is that if two sounds are completely predictably distributed in all environments, they are allophonic; if there is any degree of unpredictability, they are contrastive. Phonologists are often reluctant, however, to treat cases with a mixture of some amount of predictability and some amount of unpredictability as simply contrastive, as the examples below illustrate.

First, there are a number of instances in which phonological relationships are mostly unpredictable but in which there are a few areas of predictability. These are in fact cases of contrast neutralization; an otherwise contrastive pair is neutralized in some context or contexts. Traditionally, such cases have been assumed to still embody contrasts, but a number of researchers have claimed that such neutralization actually leads to a slightly different phonological relationship. Hume and Johnson (2003), for example, refer to neutralized contrasts as “partial contrasts” and show that partial contrasts in Mandarin Chinese tones are perceived as being more similar than fully contrastive tones. Similarly, Kager (2008), in a theoretical discussion of types of phonological relationships, also refers to contextual neutralization with the term “partial contrast,” again suggesting that contrastive relationships are not all created equal and that neutralization of a contrast changes the relationship in some fundamental way. Goldsmith (1995:11) classifies most classic cases of neutralization as cases of “modest asymmetry” on a “cline” of contrast, distinct from truly contrastive cases. Further examples are given in Hualde (2005) for Spanish and Ladd (2006) for French and Italian.

On the other hand, there are also many cases in which relationships are mostly predictable, but where there are a few environments in which the choice between sounds cannot be predicted. One example of systematic unpredictability is found in Canadian Raising, a phenomenon that has been reported for many dialects of
English, both within and outside of Canada (e.g., Joos 1942; Chambers 1973, 1975, 1989; Trudgill 1985; Vance 1987; Allen 1989; Britain 1997; Trentman 2004; Fruehwald 2007). The diphthongs [ɑɪ] and [ʌɪ] are generally predictably distributed, with [ʌɪ] occurring before tautosyllabic, tautomorphic voiceless segments and [ɑɪ] occurring elsewhere (e.g., tight [tʌɪt] but tide [tɔɪd]). There are, however, surface (near) minimal pairs distinguished by the two vowels, such as writing [raɪrɪŋ] and riding [rʌɪrɪŋ], or title [tʌɪɾl̩] and idol [aɪɾl̩], in which the two systematically contrast before a flap [ɾ].

Given the presence of such minimal pairs, it has been argued that [ɑɪ] and [ʌɪ] are contrastive in Canadian English (and other similar dialects) (see, e.g., Mielke, Armstrong, and Hume 2003), but others have been reluctant to relinquish the status of the two as allophonic, largely because the pattern is actively productive in nonsense words (e.g., Bermúdez-Otero 2003; Boersma and Pater 2007; Idsardi 2006), or because the process of flapping itself is assumed to be predictable, thus allowing the vowel quality to be predicted from the underlying representation. Other examples of systematic exceptions to otherwise predictable distributions include Bloomfield (1939: section 35) on Menomini; Dixon (1970: 93) on Gugu-Badun and Biri; Blust (1984: 424) on Rejang; Kiparsky (2003: 6) on Gothic; and Kochetov (2008: 161) on Korean.

Cases such as the ones described above illustrate the fact that the standard binary division between predictable and unpredictable distribution is insufficient to characterize the relations that hold between sounds in a language. In the next section, I present a metric that is probabilistic in nature and is designed to capture varying degrees of predictability in phonological relationships (originally described in Hall 2009).

### 3.2 A probabilistic metric

Rather than dividing predictability into two categories, predictable and unpredictable, it is quite possible to measure the degree to which two sounds are predictably distributed in a language. Consider the diagram in Figure 6.1.

In Figure 6.1, the black triangles represent some sort of categorical unit (e.g., a segment, a feature, a string, an onset ...); the circles represent the sets of environments those units can occur in. At the left-hand end of the continuum, the distributions of two sounds are entirely non-overlapping; a particular environment...
will occur in the distribution of only one of the two sounds, making it possible to predict which of two sounds will occur in that environment. At this end of the continuum, the sounds are perfectly allophonic. At the other end, the distributions of two sounds are entirely overlapping; any given environment occurs in the distributions of both sounds, making it impossible to predict which of the two sounds will occur in that environment. At this end, the sounds are perfectly contrastive. Crucially, there are an infinite number of points between these two extremes, with varying degrees of overlap; in real life, two sounds may fall at any point along this continuum. Thus while the model is designed to measure the relationship between two categorical units of phonological representation, it is a gradient model, with the gradience coming from the distances between the circles, which is a continuous measure.

In practical terms, this continuum of contrastiveness can be measured in terms of entropy, or uncertainty: given a particular environment, how uncertain are we that sound \(a\) occurs instead of sound \(b\)? If we are completely certain as to which occurs, this is analogous to perfect allophony, i.e., perfect predictability. If we are completely uncertain (i.e., we’d be making a 50-50 guess as to which of two sounds occurs), this is analogous to perfect contrast, i.e., perfect unpredictability across all environments. Any degree of uncertainty between these two extremes is associated with intermediate phonological relationships.

More explicit details of the metric can be found elsewhere (Hall 2009, 2012), so an abbreviated description is provided here. Uncertainty or entropy (abbreviated with the Greek letter H) is a measure taken from Information Theory (Shannon and Weaver 1949) and generally calculated as shown in (1a). In this equation, \(p_i\) is the probability of each possible outcome from a set of outcomes. (1b) shows a modification of this general equation to the realm of phonological relationships, where the system of interest is the relationship (or degree of contrastiveness) between exactly two elements, sound \(a\) and sound \(b\), in a particular environment, \(e\). Because of this limited case, \(H(e)\) will range from 0, in the case where the probability of one of the two sounds is 0 and the probability of the other is 1, to 1, in the case where the probability of each of the two sounds is 0.5. Thus, an entropy of 0 represents perfect allophony, and an entropy of 1 represents perfect contrast.

\[
\begin{align*}
(1) \quad & \quad a. \quad H = - \sum p_i \log_2 p_i \\
& \quad b. \quad H(e) = - \sum_{i \in \{a,b\}} p_i \log_2 p_i
\end{align*}
\]

The probability of each outcome, \(a\) and \(b\), can be calculated as shown in (2) for \(a\). This is the probability of occurrence of sound \(a\) occurring in environment \(e\), where the possibilities of sound choices are limited to \(a\) and \(b\). It is calculated by simply counting the number of occurrences of \(a\) in environment \(e\) (\(N_{ae}e\)) and dividing that number by the number of occurrences of either \(a\) or \(b\) in environment \(e\) (\(N_{ae}e + N_{be}e\)). The probability of \(b\) is calculated analogously, simply using \(N_{be}e\) as the numerator.
(2) \( p(a; a, b | e) = \frac{N_{ae}}{N_{ae} + N_{be}} \)

The number of occurrences of particular sounds in particular environments can be calculated using either types or tokens; types would be counted from a lexicon of a language, tokens from a corpus. The relative merits of each are discussed in the prior-mentioned works.

In addition to the degree of contrastiveness that exists between two sounds in any particular environment, which is calculated by (1b), it is also possible to calculate the systemic contrastiveness of a pair of sounds – how contrastive a pair of sounds is across all environments in which at least one of the sounds occurs. Systemic contrastiveness is calculated as shown in (3); the environment-specific contrastiveness \( H(e) \) from (1b) is multiplied by the probability of the particular environment, and the sum of these products is taken across all environments. Again, because of the limited choice of two outcomes, \( H \) will range from 0 to 1.

(3) \( H = \sum (H(e) \times p(e)) \)

In sum, then, the metric involves two gradient calculations: environment-specific contrastiveness, which ranges between 0 and 1 and indicates the degree of unpredictability between two sounds in a particular environment, and systemic contrastiveness, which also ranges between 0 and 1 but indicates the average degree of unpredictability between two sounds across all environments. The sounds themselves are simply countable chunks within the speech stream: conventionally, they would be segments, but any delimitable and therefore countable unit could be used as the input to this metric.

### 3.3 Applications

I now turn to a description of the utility of the probabilistic metric of phonological relationships, as described in the preceding section. Again, more detail can be found in other works; the main goal here is simply to illustrate that the addition of gradience to a model of phonological relationships has useful consequences. Three primary areas will be discussed: sound change, perception, and variation.

#### 3.3.1 Sound change

One area in which this metric is particularly useful is in documenting sound changes (see also Hall 2013b). It is well established that one way in which sound systems can change is through a change in the relationships between sounds; two sounds that are separate phonemes may merge and become allophonic variants of a single phoneme (e.g., the transition between /β/ vs. /f/ to \([v] \sim [f]\) in Old English), or two sounds that were allophonic at one stage may split and become separate phonemes at a later stage (e.g., the transition between \([v] \sim [f]\) to /v/ vs. /f/ in Modern English); see, for
example, Hock (1991). It is clearly not the case that language users abruptly shift from one kind of relationship to the other; there are intermediate stages of predictability during the transition period from one stage to another. A probabilistic metric allows us to document the change as it happens and answer questions such as:

- How far advanced is a particular change?
- Are some intermediate stages more or less stable than others?
- Can a change start to happen and then revert?
- Is there a tipping point past which a change accelerates/becomes inevitable, and if so, where is it?

An example of a possible sound change in progress that has certainly received a lot of attention and that may be illuminated by this approach is the phenomenon of Canadian Raising, introduced above in section 3.1. Hall (2012) examines this case in some detail and shows that before [ɾ], [ʌɪ] and [ɑɪ] are indeed highly unpredictably distributed, with an environment-specific entropy of 0.97, quite close to perfect contrast, but that the systemic entropy is 0.05, quite close to perfect allophony. (These calculations are made using type frequency calculations from the International Corpus of English for Canada (Newman and Columbus 2010).) Thus, the high degree of contrast in the pre-[ɾ] environment does not seem to have yet triggered a spread of unpredictability to other environments (though cf. Hall 2005 for some counterexamples); rather, the data are still almost categorically distributed, as one might expect given traditional phonological rules. At the same time, note that it is not in fact a “perfect” phonological relationship in either context; there are areas of fuzziness on both sides, which can be tracked and investigated through the use of this model.

3.3.2 Perception

It has been shown that phonological relationships affect the perception of sounds in a language (e.g., Jaeger 1980; Ohala 1982; Whalen, Best, and Irwin 1997; Kazanina, Phillips, and Idsardi 2006). Boomershine et al. (2008), for example, show that [d] and [ɾ], which are allophonic in English and contrastive in Spanish, are perceived as more similar by English speakers than by Spanish speakers, while [d] and [ð], which are contrastive in English and allophonic in Spanish, are perceived as more similar by Spanish speakers than by English speakers. As mentioned in section 3.1, Hume and Johnson (2003) show that contrasts that are neutralized are perceived as more similar than those that do not undergo neutralization. The metric described in section 3.2 suggests that in fact, there is a gradient scale of perceived similarity, with sounds that are at the low end of the entropy scale being perceived as most similar, while those at the high end are perceived as being most distinct. Hall (2009) demonstrates that this may in fact be the case with the pair of sounds [s] and [ʃ] in German, though the results did not reach significance. When these sounds were embedded in various different nonce-word contexts, they were rated as being most similar when the context was one of high uncertainty and most different when the context was
one of low uncertainty, as shown in Figure 6.2; the vertical axis shows the z-score normalized rating results, with “more similar” toward the top of the graph.

3.3.3 Variation
As a final example of the utility of including gradience in a model of phonological relationships, consider the case of inter-speaker variation. Thakur (2011) makes use of the probabilistic metric described in section 3.2 to model inter-speaker variability in Gujarati sibilant production. There is quite a bit of controversy about whether Gujarati has a contrast between [s] and [ʃ] or whether the two are allophonic or even quasi-phonemic (see Turner 1921; Grierson 1931; Pandit 1954; Adenwala 1965; Dave 1977; Masica 1991; Mistry 1997). In a controlled production study, Thakur had 20 participants produce natural Gujarati words that contained either /s/ or /ʃ/ in a variety of environments, reading from a balanced word list. She was then able to calculate the systemic contrastiveness for each individual participant, based on the set of words provided (see Figure 6.3).

![Figure 6.2](image-url)

**Figure 6.2** Relation between degree of uncertainty (environment-specific contrastiveness) and perceived similarity for the pair [s] / [ʃ] in German. “More similar” is toward the top of the graph.

![Figure 6.3](image-url)

**Figure 6.3** Systemic contrastiveness values for each of 20 Gujarati speakers in Thakur (2011). The broken line shows her division of the group into speakers with perfect or near-perfect contrast and those with partial contrast.
While 12 of her participants showed perfect or near-perfect contrast, she found that eight participants had something much closer to partial contrast, and none showed an allophonic pattern. The use of an objective metric allowed her to make further observations about the types of patterns displayed by the various speakers that would have been lost in a traditional approach; all 20 of her participants show “contrast” of some sort, and if that relationship is treated as a monolithic whole, important insights into the synchronic state of variation in a language may be lost.

4 Conclusions

There is indisputably both some degree of categoricality and some degree of gradience in phonology. Whether we think that categories emerge from, emerge with, or are primitives alongside gradient data, we can build models that incorporate aspects of both. In this chapter, I have presented one example of a metric that models phonological relationships between some sort of countable units (such as segments) in a gradient, probabilistic fashion, and shown that such a model provides insights into phenomena that, under current binary models of such relationships, would remain opaque.

Acknowledgments

I would like to thank Chuck Cairns, Eric Raimy, Beth Hume, Purnima Thakur, and the audience at the 2012 CUNY Phonology Forum for assistance with various aspects of this chapter. The usual caveats apply.

Notes

1 Ernestus (2011) actually states that there is no remnant of the alveolar gesture. It is unclear whether this means that there are in fact traces of other characteristics of the /n/ or whether she specifies the lack of an alveolar gesture simply because she is discussing place assimilation. In the original Farnetani and Busà (2004), no acoustic remnants of nasality are discussed, though such remnants are discussed for some of the other clusters examined, which may suggest that they were in fact absent in the /n/ \( \rightarrow [k] \) assimilation.

2 It should be noted that only two of their ten subjects produced such “partial” assimilations, and each of these produced only one partially assimilated token out of ten repetitions. Most other productions were categorically assimilated or non-assimilated, though they cite several other studies in which a greater amount of partial assimilation is found.

3 Idsardi, p.c., points out that such units could even be phonetically defined units, e.g., periods of voicing or aspiration within an acoustic signal.
References


Part II

What Are the Roles of Segments in Phonology?
The Opponent Principle in RcvP

Binarity in a Unary System

Harry van der Hulst

1 Introduction

In this chapter I will review the idea of representing “phonological segments” in terms of elements, namely unary building blocks which form the ultimate constituents of phonological structure. I will then defend a specific variant of this approach. Presently, unary primes are much less controversial than when the idea was first proposed as an alternative to binary features (cf. Sanders 1972; Anderson and Jones 1974) having been promoted within mainstream Generative Phonology by a number of prominent researchers (Goldsmith 1985, 1987; Sagey 1986; Clements 1985, 1992; Lombardi 1991; Steriade 1995; McCarthy 2004). Nonetheless, various approaches that have pursued this line of research for over three decades have fallen outside this “mainstream” and remain controversial, if at all acknowledged beyond a “footnote reference.” However, this chapter will not be an exhaustive overview of the various varieties of unary feature theory that are on the market. I will instead discuss two specific issues in Element Theory, which relate to whether a fourth element is needed in addition to the “core” elements |I|, |U| and |A|. This issue arises in two different ways. In one proposal, a fourth element is added which is essentially an opponent counterpart to |A|. This is the |I| (“ATR”) element proposed in Kaye, Lowenstamm, and Vergnaud (1985), and the |∀| element proposed in Radical CV Phonology (see section 3). The latter theory adopts a central principle (called the Opponent Principle) which demands that unary elements come in binary pairs. I refer to the first proposal as the 3/4 problem (I, U, A vs. I, U, A, I/∀). A different way in which the issue of adding a fourth element arises is in the form of proposals that either add a “colorless” element |ə|, called the centrality element (Anderson and Ewen 1987), or that split up the element |U| in two elements (one for “backness” |ɯ| and one for
“roundness” $|\omega|)$. I refer to both of these proposals as the $(1+2)/3$ problem (A + I, U vs. either A + I, U, $\bar{a}$ or A + I, $\omega$, $\bar{u}$). In this case, the Opponent Principle of RcvP militates against increasing the number of elements. RcvP thus ends up with four elements, forming two opponent pairs (color or place elements: I/U and aperture elements: A/∀). In addition to these four elements, RcvP also comprises two laryngeal opponent elements, $|H|$ and $|L|$, which cover a range of distinctions in the domain of tone and phonation. Various versions of Government Phonology contain similarly named elements, but there are different proposals for how these accommodate tone and phonation distinctions. In this chapter I do not discuss these laryngeal elements; see van der Hulst (to appear a, in prep. c) for their use in RcvP.

A discussion of Element Theory in a volume about phonological segments is justified because elements, unlike distinctive features, are autonomous phonological units that have “independent occurrence.” In this sense, elements are simplex segments, whereas phonemes that contain multiple elements are complex segments. Given independent occurrence, elements cannot be equated with (binary) features which are attributes of phonological segments. Elements are not attributes but rather primitive objects that can occur alone or in combinations. The issue of independent occurrence warrants more discussion since with reference to laryngeal elements, this would have to include the idea of independent occurrence on the tonal tier as autosegments. For reasons of space, I must refer the reader to van der Hulst (in prep. b and c).

2 A very brief history of Element Theory (ET)

The idea that “speech sounds” can be viewed as being composed out of smaller units (although not necessarily with stand-alone occurrence) is very old indeed and long predates the binary feature proposals of modern twentieth-century phonology. Many early works describe the articulatory mechanisms that underlie speech (from Panini to much later writers in the eighteenth century and beyond), recognizing that individual speech sounds result from an orchestrated collaboration of various articulatory actors and their movements (see Fromkin and Ladefoged 1981). A remarkable early discussion of such units can be found in Erasmus Darwin (1803) who proposed a small set of unary “elements” in terms of which, in his view, all human speech sounds could be represented. Jumping ahead to the twentieth century, it is noteworthy that Hockett (1955) also considers the use of unary building blocks. In perhaps the most important work on phonology ever written, Trubetzkoy (1939) does not propose a set of building blocks as such, although the notion of feature is certainly implied in his account of phonological oppositions, given his use of the term “merkmal” for properties of sounds. One of his classifications involves privative, equipollent and gradual oppositions, which, if translated into later paradigms, would correlate with unary, binary, and multivalued features. While Trubetzkoy’s three-way distinction suggests three kinds of features, it was Roman Jakobson, who, influenced by the developing field of Information Theory, captured all three types of oppositions in terms of binary features. This proposal, via Jakobson, Fant, and Halle
The Oponent Principle in RcvP

(1952) and Jakobson and Halle (1956), found its way into the theory proposed in Chomsky and Halle (1968). Element theory can be seen as an attempt to capture all oppositions in terms of privative primes. Modern instances of a hybrid approach, as implied by Trubetzkoy's classification of oppositions, can be found in approaches that combine the use of binary and unary features (Goldsmith 1985 and virtually all work in Feature Geometry, for example Sagey 1986; Clements 1992; McCarthy 1988).

An isolated non-hybrid approach is Sanders (1972), who proposed a “simplex feature hypothesis.” Then, Anderson and Jones (1974), in their “Three theses concerning phonological representations,” launched a new research program for Generative Phonology. The first thesis, which was would give rise to Dependency Phonology (DP), was that all phonological relations are asymmetric, reflecting a head-dependent relationship. The second thesis entailed the use of suprasegmental constituency (syllable structure, “feet” to capture stress, and beyond). The third thesis embodied the idea of intrasegmental constituency, or unary elements, organized in a set of gestures (equivalent to the class nodes of Feature Geometry). As per the first thesis, both supra- and intrasegmental structure were said to be governed by head-dependency relations. Anderson and Ewen (1987) offer a full-blown articulation of this DP program which, meanwhile, had only attracted a small following. From the beginning, DP, like Firth’s earlier prosodic approach (Firth 1948), failed to be acknowledged outside of Great Britain and, more narrowly, within developments in Generative Phonology in the United States, even though two of its central theses (suprasegmental structure, including feet, and intrasegmental “geometry”) later emerged as very influential independent developments in mainstream Generative Phonology (without any recognition of DP). The third thesis (unary primes), as mentioned in section 1, has emerged more recently in varieties of Feature Geometry, again without integration of DP results.

The core set of unary elements, proposed in DP, is formed by the “color” or “place” elements |I|, |U|, and the “aperture” or “sonority” element, |A|. This tripartite division was certainly implied in Jakobson’s work, who regarded color and sonority as the two primary axes for vowel systems (Jakobson 1941/1968). The “names” (i.e. “A,” “I” etc.) of these elements find their origin in the fact that, as phonological primes, they were first motivated in early DP work for vowels, but, from the start, the claim was that these elements generalize over both vowels and consonants. These same three elements show up as “particles” in Particle Phonology (Schane 1984, 1995). A crucial difference between DP’s use of the three primes and Schane’s particle theory is that in DP the relative “prominence” of a prime is expressed in terms of its role as either a head or a dependent in the relevant class node in the segmental structure, whereas Schane uses an additive model in which, for example, “more palatality” is expressed by multiple occurrences of the particle |I|.

Meanwhile, DP’s core properties had been re-invented in Government Phonology (GP; Kaye, Lowenstamm, and Vergnaud 1985, 1990) which, while itself hardly acknowledging DP, was (and is) equally ignored in “mainstream” phonology. Kaye, Lowenstamm, and Vergnaud (1985) present a model of intrasegmental structure which is essentially like that of DP; the notion “government” is simply the inverse of
the notion “dependency.” Relative prominence of elements is expressed, as in DP, in terms of a notion of headedness. GP did not, however, adopt the notion of grouping of elements within the segmental structure; see den Dikken and van der Hulst 1988 for a detailed comparison of both models. Both DP and GP then introduce additional elements. The DP inventory proposed in Anderson and Ewen (1987) ends up being quite rich and there has been very little development after this seminal publication with one exception (discussed in section 3). The element inventory of GP on the other hand has been subject to numerous modifications, eventually leading to a consensus among GP proponents to use a mere six elements (Kaye 2000; see Backley 2011, 2012; van der Hulst in prep. b for detailed discussions):

(1)  
   a. |A|, |I|, |U|  
   b. |ʔ|, |H|, |L|

Backley (2011) presents a particular version of this 6-element theory with many examples and motivations. He demonstrates how the elements in (1a) can be used to characterize “plain” vowels as well as place-properties of consonants. The elements in (1b) come into play when vowels have nasality, laryngeal/phonation properties, and tonal properties and to characterize laryngeal/phonation and manner properties of consonants.

The richness of phonetic coverage of each element in (1) is due to the fact that elements can be heads or dependents: a structural difference with implications for the phonetic interpretation of the elements. As pointed out in detail in van der Hulst (2011), approaches that use dependency (or government) capture the set of phonetic differences that can play a distinctive role in languages in terms of a set of primes that is significantly smaller than the set of features in binary feature theories. This is, firstly, because each element has two interpretations, depending on its role as either head or dependent. A second cause results from the fact that elements generalize over vowels and consonants. As a consequence, Element Theory formally unites phonetic distinctions for which traditional theories must use independent pairs of features, such as [+son] and [+voice], [+round] and [+labial], [+high] and [+ATR], [+high tone] and [+stiff vocal cords], and so on. This eliminates the need for arbitrary rules that state implications such as [+son] → [+voice]. In a dependency-based Element Theory, there is an affinity between voicing and sonority because both are interpretations of the same element (see Anderson and Ewen 1987; van der Hulst 1995, in prep. c for specifics). Additionally, using unary primes renders superfluous the attempts of “Radical underspecification” theorists (Archangeli 1984; Kiparsky 1982) to capture the universal asymmetry between the active and inactive poles of phonetic dimensions. For example, for the dimension of lip posture phonology only acknowledged rounding as active. This is captured in a unary approach by adopting the unary element |U|, while binary theories must include an ad hoc statement that declares [-round] to be a “default value.”

Whereas the use of the three elements in (1a) is standard in GP and DP, to add only those in (1b) is characteristic of the GP approach; DP has equivalents to these
elements (with somewhat different uses) but adds several other elements to the inventory. In (2), following Backley (2011), I provide the interpretations of these three elements, although for present purposes I have omitted some details:

\[
\begin{array}{ll}
\text{(2) Head Dependent} \\
\text{|ʔ| glottal stricture | stricture in the oral cavity} \\
\text{|H| voicelessness, aspiration | frication} \\
\text{vowels: high tone} \\
\text{|L| voicing | nasality} \\
\text{vowels: low tone}
\end{array}
\]

With these six elements, Backley describes numerous contrastive segment types and processes, claiming that the system is sufficient for representing all phonetic distinctions that are needed to capture what is traditionally called phonemic contrast in the world’s languages. Each possible phonemic segment is described in terms of a (possibly null) set of elements. One noteworthy aspect of his proposal (shared with other versions of GP) is that element expressions may be *headless* which means that no element is the head of the expression. For example, an expression consisting of the elements \(|A|\) and \(|I|\) can take three forms: \(|A \ I|\), \(|A \ I|\), \(|A \ I|\) (where underlining indicates headedness). The use of headless expressions is almost equivalent to DP’s use of expressions in which there is so-called *mutual dependency*: \(|A \Leftrightarrow I|\), \(|A \Rightarrow I|\), \(|A \Leftarrow I|\) (in this notation, the dependent is at the point of the arrow). The only difference between these two mechanisms is that the notion of mutual dependency cannot apply to a single element, whereas headlessness does. In that sense, headlessness is a more powerful device than mutual dependency. In my own approach, described in section 3, both headless expressions and mutual dependency are rejected.

Summarizing, hallmarks of both DP and GP have been the use of (a) unary primes and (b) an asymmetric relation of dependency/government. Additionally, all primes were always meant to generalize over properties of consonants and vowels. Both aspects allow an array of related phonetic interpretations for each prime. A difference between DP and GP is that DP proposes an intrasegmental grouping of the elements in terms of gestures also known as class nodes.

3 Radical CV Phonology

3.1 Basic principles

In my own work on phonological primes, I have developed an approach which takes its initial inspiration from DP (van der Hulst 1988ab, 1993, 1994ab, 1995, 1996, 2000ab, 2005, 2012a, in prep. c). Like DP, I use intrasegmental grouping of elements. However, I deviate from the specific grouping proposal in Anderson and Ewen (1987). In van der Hulst (2005), following Clements (1985), I adopt the view that each segment maximally has a tripartite structure consisting of three *classes*: [insert content here]
Harry van der Hulst

the Laryngeal, Manner, and Place class, the latter two being subclasses of the superclass, Supralaryngeal. In van der Hulst (to appear a and in prep. c), I elaborate this tripartite structure as in (3). Within each class, we find two subclasses that I call dimensions (adopting this term from Avery and Idsardi 200118), and each dimension contains two elements (which, referring to their articulatory correlates, we could call gestures):19

(3) The ‘geometry’ of phonemes in Radical cv Phonology

```
|C V| × |C V| × |C V| × |C V| × |C V| × |C V|
```

Note that I distinguish between head dimensions (dominated by a vertical line) and dependent dimensions. The internal combinatorial properties of head and dependent dimensions are not the same. Dependent dimensions do not allow combinations of elements at all (indicated by “[C×V]”); I return to RcvP-combinatorics below in more detail. The various labels for the classes are for convenience only, having no formal status in RcvP. Each unit in the structure can be defined in purely formal terms. The elements |C| and |V| are also strictly formal units. As mentioned, I assume that the limitation of the set of elements to two units per dimension can be seen as resulting from a basic principle of categorization, called the Opponent Principle.20 Assuming that each subclass in (3) correlates with a “phonetic dimension,” |C| and |V| correlate with maximally opposed phonetic categories within such a dimension. This, however, does not entail that phonemic21 contrast that refers to a given dimension must be expressed in terms of |C| versus |V|. A strictly minimal way of representing contrast will make use of the zero option. Thus, contrast for a given dimension can be expressed in terms of |C| versus zero or |V| versus zero; of course one expects that a choice between these two options comes with empirical consequences. For example, in the tonal dimension, either |V| (low tone) or |C| (high tone) may behave as the “marked” option with the other option being a “default.”22 While the elements are strictly substance-free cognitive units, they do correlate with phonetic events (or phonetic categories). In fact, we can think of elements as (subconscious) cognitive concepts that correlate with phonetic events/categories. The relation between formal units such as elements and phonetic events is referred to by terms like “phonetic interpretation” or “phonetic implementation.” Naturally, since the elements |C| and |V| occur in all six dimensions, these elements correlate with a wide variety of phonetic interpretations. In (4), I indicate some of these interpretations for the three head dimensions, mostly in very rough articulatory terms:
The exact phonetic interpretation of the elements is dependent not only on (a) which dimension they occur in, but also (b) on their status as head or dependent within the dimension (see below) and (c) whether they occur in a syllabic C-position ("onset") or a syllabic V-position ("rhyme").

In each dimension, then, the two elements form an antagonistic pair which is enforced by the Opponent Principle. The members of such a pair correlate with opposite extremes within a certain "phonetic dimension." These two members must have independent status because, unlike the values of binary features, they can sometimes be combined (and then enter in head-dependency relations). For ease of use, in many cases, I will adopt the mnemonic element names drawn from other element theories. Specifically, I will use element names of Dependency and Government Phonology, such as |A, U, L, ∀, I, H|. I do this to avoid cumbersome (although more accurate) expressions such as "|Place: V|" (= "|U|") (where the term "place" is a shorthand for a structural position in the segmental structure):

In comparing RcvP to other feature theories, I will also sometimes use labels such as [round], [ATR], and so on. Where it is relevant to remind the reader of the C- or V-nature of an element, I will write "V/U" or "V/[round]."

As mentioned, in some dimensions, elements can enter into combinations where each combination is maximally binary. This is indicated in (3) by "|C×V|." Specifically, this is needed in the head dimensions of Manner and Place. However, this level of complexity is not required in the Laryngeal head dimension (see van der Hulst, to appear a). In addition, none of the dependent dimensions requires combinations of |C| and |V|; this is indicated by "|C⊗V|." The fact that combinations are allowed in head dimensions but not in dependent dimensions is a clear instance of a head-dependent asymmetry. While dependents can never be more complex than heads, heads typically allow greater complexity than dependents (Dresher and van der Hulst 1998; Harris 1990). Thus, the Manner and Place class allow the following 12 structures:
The option of having structures that lack a head dimension element, which would create two additional possibilities (∅+C, ∅+V), is simply not available as part of the RcvP syntax (because dependents cannot be more complex than heads). As a result, elements in dependent nodes can only be present if there is an element in corresponding head nodes. As indicated in (6), the four-way distinction in the first column regards the combinations of elements within the head dimension. The second and third columns represent a combination of each of these four options and one element in the dependent dimension. The full array of structural possibilities in (6) is only exploited in the manner class (for both consonant and vowels) and in the place class for consonants; for vowels, apparently, we do not need the dependent place dimension; (see 8). The laryngeal class is the most limited class in that element combinations are excluded in both the head and the dependent dimension. As I show in van der Hulst (to appear a) the laryngeal class only needs the following subset of options:

(7)  
<table>
<thead>
<tr>
<th>C</th>
<th>C+C</th>
<th>C+V</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>CV+C</td>
<td>CV+V</td>
</tr>
<tr>
<td>VC</td>
<td>VC+C</td>
<td>VC+V</td>
</tr>
<tr>
<td>V</td>
<td>V+C</td>
<td>V+V</td>
</tr>
</tbody>
</table>

In this chapter, I cannot justify the required set of RcvP-structures, and I must refer the reader to van der Hulst (in prep. c) for a full exposition.

The possibility of combining elements within a head dimension can be seen as one way of capturing the fact that some phonetic dimensions can give rise to more than two contrastive options, forming a 4-way scale. The combination of elements can be seen as an instance of recursion as an element can be said to contain an instance of itself (or of the antagonistic element):
and $\forall$] rather than as $|\mathcal{A}, \mathcal{A}|$ and $|\mathcal{V}, \forall|$. The combination of elements within a dimension captures the discrete scale-like character of contrastive option within a phonetic dimension and results in a fixed limit on the number of categories (up to four). In (8c), I indicate, in rough terms, the interpretation of the four manner categories for vowels and obstruents respectively. The distinction between these major categories of phonemes is made in terms of syllable structure positions (see van der Hulst 1996, in prep. c). One might ask why this recursive split of phonemic categories halts after one loop. I surmise that this is due to the fact that a further corresponding subdivision of phonetic spaces would create problems for the auditory detection of the distinctions between the resulting categories.

A reduction of the set of elements to just two elements, $|\mathcal{C}|$ and $|\mathcal{V}|$, is possible because each dimension contains exactly two elements. This allows us to say that the element labels $|\mathcal{A}|$, $|\mathcal{U}|$, $|\mathcal{L}|$, and so on, because they occur under structurally different nodes, are paradigmatically speaking in complementary distribution, and thus can be reduced to one and the same element, viz. $|\mathcal{V}|$. The same holds for $|\mathcal{V}|$, $|\mathcal{I}|$, and $|\mathcal{H}|$, which can be reduced to $|\mathcal{C}|$. Complementary distribution is a familiar criterion that is used to reduce allophones to phonemes (where allophones are in complementary distribution in a syntagmatic sense). However, the same criterion can be applied to elements, provided that the elements that we reduce to either $|\mathcal{C}|$ or $|\mathcal{V}|$ have something in common. Commonality, known as phonetic similarity, again is a criterion for grouping allophones under one phoneme. In the case of elements, the commonality is that $|\mathcal{A}|$, $|\mathcal{U}|$, and $|\mathcal{L}|$ represent vowel- or rhyme-oriented choices, and so reduce to $|\mathcal{V}|$, while $|\mathcal{V}|$, $|\mathcal{I}|$, and $|\mathcal{H}|$ represent consonant- or onset-oriented choices, and reduce to $|\mathcal{C}|$ (again the choice of labels is merely for convenience). It is important to note that $|\mathcal{C}|$ and $|\mathcal{V}|$, despite their respective onset and rhyme bias, can occur in both onset and rhyme positions. For example, in the manner head dimension, $|\mathcal{A}|$ is a vowel-oriented element because in the syllable nucleus (i.e., the head of the rhyme), this element is the preferred (unmarked, optimal) choice, denoting maximal openness and sonority. On the other hand, $|\mathcal{V}|$ is a consonant-oriented element, because it is preferred in the syllable onset, where it denotes closure and hence minimal sonority.

Backley (2011) observes that his six GP elements form “antagonistic pairs” – much as in RcvP (a model that he refers to in other places in his book) although his model provides no formal basis for any such groupings by lacking the dimension class nodes. In RcvP, on the other hand, antagonistic (or opponent) grouping forms a pivotal and formal part of the theory since it expresses the idea that phonology is based on contrast.

In summary, given the anatomy of the human speech apparatus, RcvP acknowledges classes within which contrast can be expressed. Within these classes, the Opponent Principle enforces an equipollent contrast between two elements that can be multiplied within head dimensions by two using dependency relations, leading to a maximal four-way scale. The possibility of adding on a dependent dimension allows for a limited set of further distinctions.
In a sense, RcvP can be understood as a meta-theory of phonological features/elements. The Opponent Principle and the “X-bar” architecture of the phoneme predict a limited set of features/elements that, as I show in detail in van der Hulst (in prep. c), that conforms to a number of empirically well-motivated partial feature theories in the domains of tone, phonation, place, and manner.

3.2 Vowels

Ignoring the laryngeal elements (for phonation, nasality, and tonal properties), the place and manner elements of RcvP characterize vowels into 25 categories, which roughly correspond to IPA symbols:\textsuperscript{29}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
 & \textsuperscript{I} & \textsuperscript{IU} & \text{Placeless} & \textsuperscript{UI} & \text{U} \\
\hline
\text{∀+∀} & i & y & i \sim w & u & u \\
\hline
\text{∀} & i & y & ə (~ schwa) & \text{ʊ\textsuperscript{30}} & \text{ʊ} \\
\hline
\text{∀A} & e & o & ə \sim y \sim æ & \text{θ} & \text{θ} \\
\hline
\text{Δ∀} & e & æ & æ \sim æ & \theta & ə \\
\hline
\text{A} & æ & æ & a & a & a \\
\hline
\end{tabular}
\caption{}
\end{table}

We arrive at this table as follows. As mentioned, the full array of place options is only used for consonants; for vowels we only need the structure in (10a). For manner, vowels (and consonants) use the full array of structure in (6), here repeated in (10b):

\begin{enumerate}
\item[Vowel place options]
\begin{align*}
\text{C} & \quad \text{C} + \text{C} \\
\text{CV} & \quad \text{CV} + \text{C} \\
\text{VC} & \quad \text{VC} + \text{C} \\
\text{V} & \quad \text{V} + \text{C} \\
\end{align*}
\item[Vowel manner options]
\begin{align*}
(+C = \text{pharyngeal cavity, i.e. advanced}; +V = \text{nasal cavity})
\text{C} & \quad \text{C} + \text{C} \\
\text{CV} & \quad \text{CV} + \text{C} \\
\text{VC} & \quad \text{VC} + \text{C} \\
\text{V} & \quad \text{V} + \text{C} \\
\end{align*}
\end{enumerate}

By combining the manner and place options, we allow 16 structures. I have added one row for high advanced vowels that are distinguished from high non-advanced vowels by having a dependent dimension specification |∀|.\textsuperscript{31}

We must bear in mind that in RcvP, headedness is stated \textit{per dimension}. Also, I remind the reader that, unlike in GP, dimension expressions cannot be
headless; nor does RcvP acknowledge the DP option of mutual dependency. The rejection of headless or mutually dependent expressions constitutes a major difference between RcvP and DP or GP. All three approaches, however, allow the null-option (but only in the absence of a dependent). This last point raises an important issue. By allowing a 5-way distinction along the place axis (including the placeless option), we de facto admit that the absence of a dependent place specification can be contrastive with the presence of a dependent place specification. We have observed in section 3.1 that this is also true at the level of dimensions (see 6). In other words, to characterize central vowels as placeless is unproblematic.

In closing this section, I return to the point that elements have a variety of phonetic interpretations. The following table makes this explicit by focusing on the fact that the articulatory properties of the elements, can be unified in acoustic terms, that is, in terms of their effect on formant properties:

<table>
<thead>
<tr>
<th></th>
<th>Head</th>
<th>Dependent</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>H (C_{lar})</td>
<td>high register</td>
<td>high tone</td>
<td>raising F_0</td>
</tr>
<tr>
<td>L (V_{lar})</td>
<td>low register</td>
<td>low tone</td>
<td>lowering F_0</td>
</tr>
<tr>
<td>∀ (C_{man})</td>
<td>ATR-closed</td>
<td>open</td>
<td>lowering of F_1</td>
</tr>
<tr>
<td>A (V_{man})</td>
<td>RTR-open</td>
<td>closed</td>
<td>raising of F_1</td>
</tr>
<tr>
<td>I (C_{place})</td>
<td>front-spread</td>
<td>spread</td>
<td>raising of F_2</td>
</tr>
<tr>
<td>U (V_{place})</td>
<td>back-round</td>
<td>round</td>
<td>lowering of F_2</td>
</tr>
</tbody>
</table>

The laryngeal elements, |H| and |L|, determine the fundamental frequency (F_0, correlating with pitch level) of vowels, allowing for tonal distinctions. The elements |∀| and |A| determine the relative size of the pharyngeal cavity which has consequences for the oral cavity. |∀|, by advancing the tongue root, increases the pharyngeal cavity and consequently decreases the oral cavity. This lowers F_1. This effect can be phonetically enhanced by closed jaw position. The element |A|, by retracting the tongue root, does the opposite and thus raises F_1. Open jaw position enhances the effect of |A|. The color elements |I| and |U| bear on the length of the oral cavity in front of the oral stricture, which is longer for back vowels that have lower F_2. Lip rounding increases the length of this oral “tube” which is why lip rounding is said to enhance backness by lowering F_2 further. Given this view of the duality of elements, we can say that in dependent position, elements activate the enhancing mechanisms only. In this section, I have discussed the basic architecture of RcvP with specific reference to place and manner elements of vowels. We have seen that the inventory of elements reflects a cognitive principle of categorization (the Opponent Principle),
which promotes a binary polar contrast within each phonetic dimension. This principle captures the core foundation of phonology which is to achieve the optimal expression of contrast. Given that there are three classes of elements (laryngeal, manner, and place), each capturing two dimensions that allow for only two elements, all contrast can be expressed in terms of two unary elements. Interestingly, the search for phonological primes ends up with one binary pair of unary elements rather than with an arbitrary list of binary or, for that matter, unary features.

4 The 3/4 problem

When comparing the two models (current GP and RcvP), one might conclude that there is very little difference (here, for the sake of comparison, using the “convenient” element labels that have no official status in RcvP):

(12)

<table>
<thead>
<tr>
<th></th>
<th>RcvP</th>
<th>GP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manner</td>
<td>A ∀</td>
<td>A ?</td>
</tr>
<tr>
<td>Place</td>
<td>U I</td>
<td>U I</td>
</tr>
<tr>
<td>Laryngeal</td>
<td>LH</td>
<td>LH</td>
</tr>
</tbody>
</table>

However, there are some substantial differences with regard to the interpretations of the six elements in their various head and dependent roles. Referring to van der Hulst (in prep. c) for an extensive comparison, I will briefly focus on the elements | ∀ | and | ? |.

In RcvP, the element | ∀ |, like | ? | in GP, represents non-continuancy in consonants, this same element represents ATR (and vowel height) in vowels only in RcvP (see 9). The element | ? | is not so used in GP. It can be used for vowels, but then it denotes a phonatory property like laryngealization or glottalization. So, how does GP represent ATR?

Interestingly, Kaye, Lowenstamm, and Vergnaud (1985) proposed a fourth element (| ̵ |), which, like my | ∀ |, was meant to express “ATR,” but this element (which had the arbitrary restriction that it could only be used as a head) was later abandoned, being replaced precisely by the mechanism that RcvP has excluded, namely a contrast between being headed (implying [ATR]) and being headless; I call this contrastive use of headedness.35 For example, in current GP (including Backley’s 2011 version), the difference between /i/ and /ɪ/ is that the former vowel has a headed element, e.g. | ̵ |, while the latter is non-headed | ̵ |. One might conclude that not much is at stake here (a trade-off between an extra element or allowing headless expressions). However, it should be clear that the extra element, | ∀ |, is “predicted” by the “contrastive logic” (i.e., the Opponent Principle) of RcvP: each
dimension contains two antagonistic elements. Thus, there must be a counterpart to \(|A|\) in the manner dimensions. Given that this is so, a possibility presents itself to avoid headless expressions, which are at odds with the foundational assumption of a “pure” dependency approach that only combinations of units are characterized by an asymmetric relationship of dependency. From this “first principle,” it follows that neither headless expression (nor expressions showing “mutual dependency”) should be considered.

Although all versions of Element Theory contain elements in addition to the core set \(|IUA|\), it could be argued that RcvP (as the original GP theory) puts forward a fundamental change of the element approach. Clearly, additional elements are needed for distinctive properties involving nasality, tone, phonation, and perhaps for various types of consonantal stricture. However, the addition of \(|\forall|\), as a counterpart to \(|A|\), just for basic vowels, presents a complication of the core set, which derived much of its appeal from giving a straightforward explanation for the “triangular nature” of the majority of vowel systems that, generally, show less vowel contrast in the lower regions of the vowel space than in the upper region. This being said, even DP did not confine itself to the basic IUA-set when more complex vowel systems were considered. In fact, Anderson and Ewen (1987) add both an ATR element and a “centrality” element to their system, whereas Kaye, Lowenstamm, and Vergnaud (1985), as shown, adopted their \(|I|\) element, as well as a special element called the “cold vowel” which shared some characteristics with DP’s centrality element.

In the next section, I will suggest that there are, in fact, reasons for supporting a triangular element system, but that there are also reasons for supporting a quadrangular system. The “pressure” for “3” comes, as I will show, from properties of the human articulatory apparatus, which, when considering the available muscular activities of the tongue, suggest three constriction loci. I will base this claim on the work by the phoneticians Sidney Wood and Ken Stevens. The pressure for “4,” on the other hand, comes from a cognitive force that I have called the Opponent Principle, which captures the idea that the phonetic resources for speech are categorized in a system of opponent elements that correlate with maximally dispersed acoustic events. My claim is that this cognitive force, which provides a raison d'être for phonology as being distinct from phonetics, is responsible for “creating” the fourth element (namely \(|\forall|\)) which functions as the counterpart to \(|A|\). As I will argue, the element \(|\forall|\), in need of an articulatory basis, either draws on the same articulatory resources that also underlie the element \(|I|\) or correlates with the overall tenseness of the articulation.

5 Wood’s system of articulatory features

The phonetician Sidney Wood (e.g., Wood 1979, 1982, 1990) recognizes four constriction locations for vowels, three of which are taken to be basic and thus deserving of their own distinctive feature. I summarize Wood’s views on constrictions
locations in the following table, IPA symbols, features, and articulatory and/or acoustic correlates:

(13)

<table>
<thead>
<tr>
<th>Constriction locations</th>
<th>Muscle activity</th>
<th>Vowels</th>
<th>Features</th>
<th>Phonetic effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palatal</td>
<td>genioglossal activity</td>
<td>[i-ɛ, y-œ]</td>
<td>[palatal]</td>
<td>widens the lower pharynx, raises the tongue body</td>
</tr>
<tr>
<td>Palato-velar</td>
<td>styloglossal genioglossus palatoglossal</td>
<td>[u, ʊ, ů]</td>
<td>[velar] [palatal]</td>
<td>draws tongue toward nasopharynx, widens lower pharynx (F1 below 350Hz), locates palate-velar constriction precisely, F2 beyond 1250 toward 1500</td>
</tr>
<tr>
<td>Pharyngovelar (i.e., Uvular)</td>
<td>styloglossal superior pharyngeal constrictors</td>
<td>[o, ɔ, ɤ]</td>
<td>[velar] [pharyngeal]</td>
<td>F2 about 800 Hz for rounded [o]-like vowels</td>
</tr>
<tr>
<td>Pharyngeal</td>
<td>hyoglossal and/or superior and middle pharyngeal constrictors activity</td>
<td>[a, æ]</td>
<td>[pharyngeal]</td>
<td>narrow lower pharynx, raises F1 beyond 600 Hz</td>
</tr>
</tbody>
</table>

Wood also proposes two additional features:36

(14) a. **[open]** refers to lower mandible position  
     b. **[tense]** refers to:  
        (i) increased activity in the lingual musculature for the bunched tongue position ([i,e,u,o] vs. [i, ɛ, ʊ, ɔ]),  
        (ii) more laryngeal depression, and;  
        (iii) increased labial activity for rounded vowels [u, o] vs. [ʊ, ɔ]
The four constriction loci correspond with the following vowels (among others; see 13, third column):

(15) [i] [u] – [o] [a]

The “missing” [e] type vowel would result from combining the “[i] constriction” with an open jaw position.

As indicated in the table in (13), Wood proposes three phonological features which are very similar to the AIU elements. (As he points out, this three-way distinction is already known from old Indian linguistic traditions.) Each feature correlates with a separate muscle group:

(16) Palatal: genioglossus (~ |I|)  
Velar: styloglossus (~ |U|)  
Pharyngeal: hyoglossus (~ |A|)

The following drawing shows the three major muscle bundles:

(17) Genioglossus  Styloglossus  Hyoglossus

I conclude that Wood’s theory supports the triangular IUA system from a muscular-articulatory point of view.

One point of difference, however, is that in Wood’s system, high non-front vowels are specified as [+palatal]. This is where the correspondence between his feature [palatal] and the element |I| breaks down:
However, we can adopt the viewpoint that the element |U| does indeed draw on genioglossal activity. This in itself explains the often observed affinity between |I| and |U| (cf. Ewen and van der Hulst 1988). Thus, rather than linking elements to muscular activity in a one-to-one fashion, I suggest that both |I| and |U| draw on the genioglossal activity; the former more so than the latter. In a sense, this activity characterizes the place node itself. That elements can draw on more than one muscular activity is shown by the fact that Wood makes reference to the genioglossus and palatoglossal activity (which elevates the posterior part of the tongue) for his feature [palatal] when it characterizes high non-front vowels ([u, ʊ, ɯ]). Also, as shown in (13), the feature [pharyngeal] (i.e., element |A|) also draws on different muscular activities (i.e., hyoglossal activity and pharyngeal constrictors). It would seem though that each feature/element corresponds to a primary muscle group but can also draw on a secondary group.

An obvious criticism of triangular systems has always been that there is no feature/element that captures the natural class of high vowels. Nor does this set express the property of ATR (or “tense”), which is why DP and GP added an ATR element (which GP later, as we have discussed, replaced by contrastive headedness). The muscular bundle that could be held responsible for advancing the tongue, and thus raising it, is the genioglossus, which is already held responsible for creating a palatal constriction. A crucial point is that RcvP, given its design, enforces a fourth element which can capture both height and ATR. Beyond that, analyses of height and ATR vowel harmony systems support the need for this element (see van der Hulst 2012a, b, c, to appear b and, especially, in prep. a). The important question is now how one can motivate this fourth element in terms of its articulatory mechanism and corresponding acoustic effects? We must first bear in mind that the fourth element can occur in both the head manner component and in the dependent manner components (see 3). When this element occurs in the head aperture component, its interpretation can best be described in terms of vowel height or aperture.

When occurring in the dependent manner component, I suggest that there are, in fact, two ways in which the fourth element can be phonetically grounded and that these two ways might reflect a pattern of variation among languages that appears to be real. Ewen and van der Hulst (2001) point out that languages seem to be complementary in using either a tense/lax (or peripheral/central) distinction or a [+ATR]/[-ATR] distinction in their vowel system. Here I propose that these different phonetic mechanisms are complementary (in the sense that no language uses both) because...
they are manifestations of the same element, namely $\forall$ and, moreover, that the ambiguity in the phonetics of this element is caused by the fact that it is not grounded in its own unique primary muscle group and must therefore draw either on a muscle group that is active as the primary group for another elements or on the additional phonetic dimension of overall muscle tension.

To begin with the latter option, the most obvious articulatory correlate of the fourth element $\forall$, from the viewpoint of Wood’s theory, would be the feature [Tense]. In that case, $\forall$ would correlate with increased overall muscular activity (relative to specified constriction) and some additional properties (see 14b). So, even though the phonetic reality of the feature [tense] has been called into question (Fischer-Jørgensen and Jørgensen 1969), I accept Wood’s finding that there is, in fact, an observable set of articulatory correlates for this feature/element. The former option for the fourth element $\forall$ leads to what has been called [Advanced Tongue Root] (which has often been proposed as a substitute for [tense] as a phonological prime). In this case $\forall$ draws on the genioglossus which is also the primary muscle for the element [I]. It is noteworthy that the element [I] (and Wood’s feature [palatal]) correlates with a tongue maneuver that causes both a palatal constriction and an advancement of the tongue root. My proposal here is that the phonetic properties [ATR] and [tense] are both possible interpretations of the element $\forall$.

In conclusion, my suggestion is that the fourth element (as it occurs in the dependent aperture component) either draws on the same phonetic substance that forms the basis for the element [I] or captures the phonetic property of tense which acts as an “operator” on all constrictions:43

Summarizing, my suggestion is that the binary categorization of phonetic dimensions that underlies the architecture of RcvP is rooted in a cognitive principle: the Opponent Principle. Given the biological availability of our articulatory mechanisms, discrepancies may arise between the demands of cognitive systems and the anatomy on which these systems are “superimposed.” I have suggested that, as a result, the cognitive category $\forall$ (as it occurs in the dependent aperture component) draws on two types of phonetic substances – one of which involves “double dipping.” When this same element occurs in the head aperture component, its interpretation can best be described in terms of vowel height or aperture (in which case many descriptions will also often use the term pair “tense/lax”).

(20) Genioglossus Styloglossus Hyoglossus

\[\text{I}_{\text{Place}:C} \quad \text{U}_{\text{Place}:V} \quad \forall_{\text{Man}:C} \quad \text{A}_{\text{Man}:V} \]

\[\text{Tense}\]

---

*The Opponent Principle in RcvP* 165
6 Articulatory and acoustic correlates of elements

Wood’s theory of features is based on articulatory mechanisms. In parallel with his theory, Ken Stevens (Stevens 1972) developed a theory about a specific nonlinear relation between articulatory mechanisms and acoustic effects which shows that the latter can be relatively insensitive to small changes in certain constriction areas (see also Stevens and Keyser 2010). The theory is called quantal, because when small changes along an articulatory area pass a certain threshold there is a clear acoustic effect which corresponds with a feature change (or feature value change). In terms of constriction loci in the vocal tract, Stevens’ quantal areas correspond exactly to Wood’s three features: [palatal], [velar], and [pharyngeal]. Given Wood’s explicit account of these three loci in terms of muscle groups, it would seem that a quantal effect emerges when a different muscle group is activated or becomes “dominant.” It would seem that there is a straightforward correlation between articulatory mechanism and stable acoustic effects.

In GP (as well as in DP), it is usually claimed that the three elements, |I|, |U|, and |A|, correlate with acoustic images in the mind of language users. Backley (2011) places articulatory correlates outside the grammar. This viewpoint can be traced back to Roman Jakobson’s idea that since the acoustic aspect of speech is shared by speaker and hearer (existing “in between them” so to speak), acoustics must have primacy over articulation. An additional argument that is often made is that articulation is highly variable and that speakers can reach the desired acoustic targets in different ways even when having “a mouth full of marbles.” But the issue remains controversial since consonants of different places of articulation can hardly be said to have invariant acoustic properties, given that their identity is revealed by formant properties of following vowels (e.g., see Delattre, Liberman, and Cooper 1962). While acknowledging that speakers can adapt articulations in special circumstances to reach their acoustic goals, Taylor (2006) argues, convincingly in my mind, that phonemes must be associated with specific articulatory plans, which, perhaps more for consonants than for vowels, represent what is constant in phonetic events that cannot be easily unified in terms of their acoustic properties. In the so-called motor theory of speech perception (Liberman and Mattingly 1985) it is even claimed that we perceive acoustic events in terms of motor representations that cause such acoustic events. This is the view that has been integrated with theories about mirror neurons (Fowler and Galantucci 2002). To resolve the debate about whether articulation or psycho-acoustics is primary, I suggest that, while all phonemes are represented in terms of elements that correspond to both an articulatory plan and an acoustic image, the dominance of these two aspects differs for vowels and consonants (yet another instance of a head-dependency relation):

\[
\begin{array}{c|c}
\text{Vowels} & \text{Consonants} \\
\hline
\text{Acoustic image} & \text{Articulatory plan} \\
\text{Articulatory plan} & \text{Acoustic image}
\end{array}
\]
After all, in vowels, the articulatory plan is necessarily “vaguer” because the actual target of articulation cannot be contacted. In consonants, on the other hand, acoustic properties are less clearly identifiable (especially for obstruents), which suggests that for these types of segments, the articulatory plan must be the unifying factor.

I conclude that there is no need to ban articulation from the grammar, but rather that both acoustics and articulation form necessary parts of the interpretation (i.e., “meaning”) of phonological elements. Whether phonetic interpretation forms part of the phonological grammar is largely a terminological issue. While the computational aspect of phonology need only make reference to structures and element, and thus not to the phonetic correlates of either, a full account of phonology must also include how the formal phonological expression correlates with the production and the perception system (see van der Hulst, to appear b).

7 The 2/3 problem

I now turn to another controversy where RcvP takes a stand that also directly follows from the Opponent Principle. In various versions of Element Theory, there have been certain problems surrounding the element |U|. First, it has been claimed that |U| can only occur in one combination with |I| (essentially leading to removing dependency as a necessary relation in the combination of these two elements). A second idea, sometimes connected to the first one, has been to add a new element, which in DP is called the centrality element |ə| (an “anti-color” element), to represent central unrounded vowels. Thirdly, it has been proposed to replace the element |U| by two elements: one for backness and one for roundness. I will first discuss the restriction on |I,U| combinations and argue that this restriction is not only theoretically arbitrary, but also undesirable empirically. The second and third ideas have in common that an extra element is added to the color class, which leads to three color elements (or two color elements and one “anti-color” element). This goes against the RcvP premise that each dimension contains only two elements, which is why this section is said to address a “2/3 problem.” With regard to the “3/4 problem,” as we have seen, RcvP must adopt the “4” choice (giving us two elements in the manner and place dimensions). In the 2/3 problem, RcvP must adopt the “2” choice to maintain the claim that the place dimensions contain two elements. I will discuss the various proposals mentioned above in turn.

Anderson and Ewen (1987: 275) propose to add a separate element |ə|, “centrality” or “non-peripherality.” The reason for this lies in the following. Although the system of DP would in principle allow for two combinations of |I| and |U|, with either one or the other as the head, Anderson and Ewen (1987: 275) state that “in virtually all languages, we find at each height maximally one segment containing both |i| and |u|; in other words, dependency relationships holding between |i| and |u| are not required.”45,46 In their discussion of the representation of central or back unrounded vowels, Anderson and Ewen, reluctant to represent such vowels as colorless (a reserve option reserve for the schwa vowel [ə]), propose to add a centrality element |ə|. This
element, as either a head or dependent in combinations with |U|, allows Anderson and Ewen to represent both central unrounded vowels and central rounded vowels (see Anderson and Ewen 1987: 224–228 for details).

In the RcvP system, different combinations of |I| and |U|, which are an undeniable theoretical option, are used to account for non-front (central) rounded vowels (often called inrounded vowels), whereas central vowels are represented as colorless; see (8). In the RcvP system, there is thus no “ban” on allowing two possible combinations for the color elements. First, such a ban is theoretically ad hoc, and secondly, the statement that no language requires both combinations (IU and IU) at a given height is incorrect if there are languages that, in addition to a front-unrounded vowel and a back rounded vowel, have two additional rounded vowels, called rounded and inrounded. A case in point is Swedish, which has a well-known contrast between [u] and [u] (see Riad 2013). Anderson and Ewen do not deny that inrounded vowels exist, but they choose to represent them in terms of the combination |uə|. Given a choice between adding an extra element and allowing two combinations for |I| and |U|, it seems obvious that there is insufficient ground for adopting the centrality element.47

Next, I will discuss the idea to replace |U| by two new elements. In various versions of Element Theory, the dual character of |U| (capturing backness and roundness, just like the feature specification [+grave] in Jakobson, Fant, and Halle 1962 or the feature [peripheral] in Rice 1995) has been called into question. A number of phonologists, notably Lass (1984: 278ff.), Rennison (1990: 187), Ritter (1997: 346), and Scheer (2004: 47ff.), have argued that these two aspects of |U| should in fact be given independent status, thus splitting up |U| into two elements, here in Lass’ symbols: |ω| (“labiality” or “roundness”) and |ɯ| (“velarity” or “high backness”). These various authors have provided different motivations for this proposal. Lass responds to an earlier version of the DP (in Ewen 1980) where central or back unrounded vowels would be represented as colorless, which he finds unsatisfactory because “all the other classes have positively specified content” (Lass 1984: 278). His proposal forces him to stipulate that the element |ω| can only occur as a dependent.

As Scheer (2004: 47ff.) points out, an important argument in this debate regards the characterization of velar consonants. Operating under the common DP/GP assumption that consonants and vowels share the same set of elements, we need to deal with the fact that in consonantal place, labiality and velarity are clearly independently needed properties, which seems to require the elements |ω| and |ɯ|. Additionally, as Scheer shows, velars induce a [u] allomorph in Czech vocative formation, [i] occurring after palatal consonants, whereas labials and dentals select the default choice of [u]. Space limitations prevent me from discussing the RcvP account of consonantal place (see van der Hulst, in prep. c). In (21), I provide the characterization of places in terms of the place elements for non-continuant obstruents:

(22) Consonantal place (stops)

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>IU</th>
<th>Placeless</th>
<th>UI</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ψ⁴⁸</td>
<td>/t/</td>
<td>/c/</td>
<td>/ʔ/</td>
<td>/k/</td>
<td>/p/</td>
</tr>
</tbody>
</table>
The phonetic interpretation of the |U| as a head is [grave] or [peripheral], shared by velars and labials, to which the element |I| adds [lingual] to characterize velars. Given these representations, the generalization that can be stated for Czech vocatives is that only these consonants that have a complex place specification can dispense their head property to the suffixal vowel which delivers [i] after palatals and [u] after velars. There are, to be sure, other empirical domains that need to be visited. Here I merely suggest an alternative to one of Scheer’s empirical arguments for splitting up the element |U|.

While the splitting up of |U| is rejected in RcvP, due to the Opponent Principle, which does not allow three color elements, Anderson and Ewen (1987) provide an additional argument. The proposal to split up |U|, in spite of making a representation of back, unrounded vowels possible without the use of a centrality element, is undesirable since it forces one to give up a direct relationship between “markedness” (in the sense of frequency of occurrence) and formal complexity which is adequately reflected by the standard DP system. That this is so follows straightforwardly from a comparison of the standard DP representations of a high back rounded vowel and a high back unrounded vowel with those of Lass (1984), given in (23).

(23) The representation of /u/: The representation of /uu/:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/u/</td>
<td></td>
<td>u</td>
</tr>
<tr>
<td>/ɯ/</td>
<td></td>
<td>u, i, ə</td>
</tr>
</tbody>
</table>

Thus, whereas in the standard DP system /ɯ/ is formally more complex than /u/, this situation is reversed in Lass’s (1984) feature system. Since it is generally assumed that a high back vowel that is rounded is less marked than an unrounded one, Lass’s (1984) system clearly does not mirror markedness (as Lass himself also explicitly acknowledges saying that all markedness consideration should be excluded from phonological representations).

However, the correlation between markedness and complexity is deserving of some further discussion. A virtue of unary systems is no doubt that an account of markedness needs much less additional machinery in the form of underspecification, marking conventions and default rules compared to binary theories. Less marked segments contain fewer elements than more marked segments. However, there is one wrinkle in the correlation. RcvP treats unrounded central vowels, such as [ɨ] or [ɯ] (which I take to never be in contrast; see 9) as colorless, which seems to imply that such vowels are less marked than vowels that contain the elements |I| and/or |U|. This is here illustrated with the high vowel row taken from (9):

(24) High vowels

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>IU</th>
<th>Colorless</th>
<th>UI</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>∀ ( ∀ ⊕ )</td>
<td>i</td>
<td>y</td>
<td>i ~ u</td>
<td>u</td>
<td>u</td>
</tr>
</tbody>
</table>

There are two points to be made here. First, by adopting the combination |UI| as the representation for [u] there is no complexity difference between this rather rare
vowel and the also rare, but more common vowel [y]. This shows that there are limits on correlations between complexity and markedness. In this case, we have to be satisfied that both [u] and [y] are more complex than the common [i] and [u]. Second, assuming that markedness correlates with complexity, it would seem to follow that non-front-unrounded vowels [i ~ uu] would have to be the most common vowels. However, I suggest that complexity is not an actual fundamental determinant of markedness. Rather, what makes segments unmarked is to have a perceptually clear and salient identification, which, I submit, is achieved by being characterized in terms of precisely one color element. This is why [i] and [u] are less marked than [y] and [uu]. It might now be asked why [uu] is more marked than [a] and, also, whether [a] is less or more marked than [i] and [u]:

(25)  
\[
\begin{array}{cccc}
V_C & V_C & V_C & \Lambda_V \\
I_C & \_ & U_V \\
\end{array}
\]

In RcvP, [a] is less marked than [uu] because the |A| element is more preferred for vowels than the |V| element. This is revealed by realizing that the former is a V-element, while the latter is a C-element (see 4 or 11). As for [i] and [u] vs. [a], we have conflicting factors. In its manner element, [a] has the preferred V-element. On the other hand, [i] and [u] have a color identification which [a] lacks. Result: a draw. Finally, comparing [i] and [u], [u] would have to be less marked, because its color element is a V-element. This is seemingly in contradiction to [i] – perhaps because it is a more frequent epenthetic vowel. However, being preferred as an epenthetic vowel does mean being a more preferred vowel. Rather, the contrary is the case: to fill epenthetic slots, languages use the least preferred vowels, namely vowels that “sneak in” precisely because they are not very salient.

In conclusion, I have here defended the (what I would refer to as the original) 2-theory of color against two versions of the newer 3-theory, one with an extra centrality element and the other with a dual substitute for |U|. The 2-theory follows from the overall architecture of RcvP as demanded by the Opponent Principle. Second, it is theoretically preferred in not having to stipulate arbitrary restrictions on element combinations or the exclusive occurrence of elements as either heads or dependents. Thirdly, it is more consistent with the idea that phonological representations give expression to markedness. Finally, as demonstrated in van der Hulst (2011, 2012abc, to appear b, in prep. a) in an extensive study of vowel harmony systems, adoption of the element |V| is also empirically motivated.

8 Conclusions

In this chapter, I have defended the architecture of RcvP (with specific reference to vowel representations), which is governed by a cognitive principle that favors polar opposites, the Opponent Principle. This principle has a perceptional rational in
The Oponent Principle in RcvP

In terms of categorical perception (Harnad 1987) and maximal dispersion (Liljencrans and Lindblom 1972). It can perhaps also be motivated in terms of the neurophysiology of the brain. The relevant principle is that polar opposites within phonetic dimensions form the optimal phonemic contrast. This principle give rise to systems of primes that are not arbitrary lists (as exemplified by traditional feature theories and all other versions of Element Theory), but rather to systems that contain opponent opposites, which may, as argued, even lead to primes that are not straightforwardly motivated by a unique or obvious phonetic correlate, namely the element $|\forall|$. I referred to the problem of using three elements (I, U, A) or four (I/U and A/\forall) as the 3/4 problem. I discussed in detail referring to Wood’s theory of articulatory features – how the fourth element shares articulatory resources with the element $|I|$. I then turned to a second problem, the 2/3 problem, which regards a debate between theories that either add a centrality element or split the element $|U|$ in two separate elements. Either proposal increases the set of color elements from 2 to 3. Here, I argued against such an enrichment to three elements by showing it to be theoretically undesirable and empirically unmotivated, concluding that we can limit the set of color elements to 2. The overall conclusion is, then, that only four elements are required to represent the place and manner properties of vowels (excluding tone and phonation, which require an additional polar pair of elements, $|H|$ and $|L|$). With regard to the phonetic interpretation of the elements, two points were made. First, given the fact that elements can be heads or dependents, elements can correlate with different articulatory interpretations that are unified in terms of their acoustic effects. Second, it was suggested that, while vowels and consonants share the same set of elements, acoustic interpretation may be more important for the former, whereas the latter are unified in articulatory terms.

Acknowledgments

I would like to thank the editors of this volume for their helpful suggestions on an earlier version of this chapter.

Notes

1 For more information, I refer the reader to den Dikken and van der Hulst (1988) as well as Backley (2011, 2012) and van der Hulst (in prep. b).
2 Kaye, Lowenstamm, and Vergnaud (1985) also had another extra element, called the “cold vowel,” which shares some properties with the centrality element of Anderson and Ewen (1987); see den Dikken and van der Hulst (1988) for discussion. This element was later abandoned and central vowels came to be represented as colorless or in terms of empty skeletal positions.
3 A similar “two place” model is proposed in Rice (1995).
4 As reported in Ohala (2004).
5 Proposals have also been made to generalize multi-valued features; see Williamson (1977) and Gnanadesikan (1997).
John Anderson in much subsequent work has shown that asymmetry is a characteristic of all linguistic structure; see especially Anderson (2011) for an extensive (three-volume) review of his work in phonology, morphology, and syntax.

The term “gesture” as equivalent to “class node” is unfortunate. The term “gesture” as used in Articulatory Phonology (Browman and Goldstein 1986) is rather equivalent to the unary elements (called components in DP) themselves.

The use of dependency structure among syllables to represent stress was later proposed in Liberman and Prince (1977), giving rise to metrical phonology. The use of groupings of elements within segments, which was discussed in van der Hulst and Smith (1982), prefigures the class node idea of Feature Geometry proposed in Clements (1985) and Sagey (1986).

As is shown in Den Dikken and van der Hulst (1988), the four unary “major articulator” features of Feature Geometry ([labial], [coronal], and [pharyngeal]) as very similar to the unary elements [U, I, A], with the fourth articulator [dorsal], shared usage with “fourth” element (whether [a], [v], [l], or [v]).

Donegan (1978) also recognizes palatality, labiality (color properties), and retraction (a sonority property) as the three basic vowel ingredients but still uses binary features in the formal representation of segments and processes or rules.

This recaptured the original claim of Jakobson, Fant, and Halle (1952) who also proposed a unified set of features for consonants and vowels. Chomsky and Halle (1968) adopted largely distinct feature sets for these two major categories. Clements (1992) then also recaptured the idea of a unified set within the Feature Geometry approach, capturing, without recognition, the unified approach of DP.

I first learned about this proposal from a presentation by Jean-Roger Vergnaud at a GLOW workshop in Paris in 1982.

For intrasegmental structure (Kaye, Lowenstamm, and Vergnaud 1985) use the terms “kernel” and “operator” instead of “head” and “dependent.” Also, an early idea that elements are feature bundles which account for their phonetic interpretation as stand-alone units and in combination with other elements was later abandoned.

Anderson (2011, volume III) provides Anderson’s update of Dependency Phonology, including a discussion of RcvP.

Kaye (1988) demonstrates that the unary approach should always be explored first, since, unlike binary approaches, it can actually be falsified.

I describe Backley’s system in more detail in van der Hulst (in prep. b) where I compare it to my own system, which I discuss in section 3.

Implicit here is the claim that phonological theory does not need to supply vocabulary to express detailed phonetic properties that play no distinctive role in the languages of the world.

Avery and Idsardi (2001) propose a theory of features which also introduces the notion of antagonistic pairs, referring to Sherrington (1947) who claimed that muscles are organized in antagonistic pairs. In their theory (unlike in RcvP) members of a pair cannot both be active in a single segment nor can both be distinctive in a single language. For a comparison of this theory, called Dimension Theory, to RcvP, I refer to van der Hulst (in prep. c).

The RcvP geometry has an “X-bar”-like organization. In van der Hulst (in prep. c), I speculate that this particular organization, which appears to be shared between (pre-merge versions of) syntax and phonology in which heads can have two types of dependents (“complements” and “specifiers”) is perhaps not accidental.
A question that could be asked is why the Opponent Principle (or an extended version of it) does not enforce four phonetic spaces rather than three. I discuss this matter in van der Hulst (in prep. c) where I consider alternative segmental structures. It is noteworthy that Anderson and Ewen (1987) and other proponents of DP did propose four gestures; see den Dikken and van der Hulst (1988) for a review. In (3), I use the terms “place” for “color elements” and “manner” for aperture or sonority elements. These labels, which I use here interchangeably, have no theoretical status since each class node has a unique structural definition.

Since I use the term “phonological” as comprising both the study of contrastive or distinctive units at the cognitive level and of phonetic categories (as well as the relation between them), I will refer to the level of cognitive (“symbolic” or “formal”) representations as “phonemic.”

This is reminiscent of Radical Underspecification Theory and, indeed, there are a few cases, especially in the laryngeal class, in which elements theory uses two opponent elements that correspond to the use of the plus and minus of a binary feature such as [±high tone] or [±voice]; see van der Hulst (to appear a).

I must refer to van der Hulst (in prep. c) for an RcvP account of syllable structure and of the segment-syllable connection (see van der Hulst 1996) for an early account.

The idea that within a class, the head dimension elements must be activated before we get to the dependent elements is analogous to the fact that in vowel systems, the manner class (more specifically its head dimension which accounts for aperture) must be activated before we get to the place dimension elements. It has been shown in typological studies of vowel systems that a minimal system would use only manner (i.e., aperture), leading to a so-called vertical vowel system found in some Northwest Caucasian languages (Kabardian, Adyghe); see Lass (1984). There are no vowel systems that only use place distinctions. This further motivates the head-status of the manner class (which expresses aperture for vowels and stricture for consonants).

This correlates with the fact that universally there are many more consonant distinctions than vowel distinctions, which correlate with the greater role that consonants play in lexical phonemic contrast. This asymmetry is paradoxical since vowels are heads of syllables.

Both laryngeal and place are dependent classes, but the place class is included in the super class, supralaryngeal. Thus, the fact that the place class allows more structures than the laryngeal class is, once more, an example of an expected head-dependent asymmetry.

Salting (2005) proposes a model, “the nested subregister model,” which also represents phonological categories in terms of a double split. He applies this to vowel height and place categories and discusses the parallels of his model to RcvP.

Where different IPA symbols are placed within a single cell, the claim is that the corresponding phonetic differences do not occur contrastively in any language. Needless to say, the proper placement of vowels in specific languages in cells cannot depend on what kind of IPA symbols we use for them, but rather on the way in which these vowels function in the phonological system (systems of contrasts, phonotactic distribution, and rules). The chart in (9) deviates slightly from that in van der Hulst (2012a). I discuss alternatives in van der Hulst (to appear b).

I added this symbol, which does not occur in the IPA chart.
The expression of ATR in terms of a dependent dimension element |∀| allows a second structure for the three lower rows in (9). I discuss the use of these structures in van der Hulst (in prep c) where usually I argue that such additional structures, while formally available, are excluded by a constraint that bars the dependent |∀| for segments that contain |A|.

This is why a vowel can only be mannerless when it is also placeless. This delivers the empty vowel, often realized as a “schwa.”

For laryngeal elements, RcvP acknowledges the distinction between register and tone proper proposed in Yip (1980).

For the notion of enhancement see Stevens and Keyser (1989, 2010) and Stevens, Keyser and Kawasaki (1986). In van der Hulst (to appear a), I argue that phonological enhancement results from adding an element in a dependent dimension that is identical to the head element.

See also Ritter (1997) for a proposal to extend the use of contrastive headedness to consonantal expressions.

It would seem obvious that Wood also needs to recognize the feature [round], although he does not mention this feature in the sources that I consulted.


See Ewen and van der Hulst (1988) who propose the “higher order” element |Y| dominates both |I| and |U| to capture this point.

One might say that here we seem to be dealing with a head-dependency and an enhancing phenomenon at the muscular level, i.e., below the level of elements.

See Halle (1983) for another account of the complex relation between features and articulatory mechanisms.

This is why Ewen and van der Hulst (1988) proposed the extra element |Y|, which dominated |I| and |U|.

That there is a close affinity between advanced tongue root and palatality is shown, for example, in Turkana vowel harmony where the palatal glide /j/ induces [+ATR] on neighboring vowels; see Dimmendaal (1983) and van der Hulst and Smith (1986).

This means that in some way I was on the right track when I suggested in van der Hulst (1988a) that “front” and “ATR” are phonetic manifestations of the element |I|.

See Harris and Lindsey (1995). These authors also suggest that acoustics must have primacy because in acquisition, perception occurs before production. However, this fact does not necessarily entail that babies do not know the correlation between acoustic events and articulation, a correlation that they “practice” in their babbling stage. The motor theory of speech perception implies that this practice can perhaps be mind-internal (given mirror neurons) and even happen for individuals who grow up understanding language, while never producing it.

Anderson and Ewen’s notation for element names uses small case letters.

GP also rejected two possible combinations of the element |I| and |U|.

Granted, the difference between inrounded and outrounded vowels is typologically rare.

This element, which correlates with ATR and closure in the vowel sphere, denotes [-continuant] in the consonantal sphere. The commonality here is relative closure; see (8).

Here I ignore the dependency relations that they provide; see Anderson and Ewen (1987: 227).

The dependent dimension element |∀| need not be specified if a language does not have a contrast between two series of high vowels.
Whether phonological representations should reflect markedness remains, however, an open question; see Newmeyer (2005).

In support of this rather speculative claim, I note that the 3/4 problem is reminiscent of an issue raised by two seemingly incompatible theories of color perception that were proposed in the nineteenth century. The trichromatic theory of Thomas Young and Hermann von Helmholtz claimed that color perception is based on three basic colors for which we have specialized cells (cones) in the retina (red, green, and blue). Edward Herring proposed another theory, called the opponent theory, suggesting that visual perception is driven by two “perceptional modules” (red/green and blue/yellow). The difference is that while red and green are paired, for the second pair a fourth “color,” namely yellow, has to be added to the model. The contradiction between these two theories was resolved when it was established that the Young/Helmholtz-theory reflect the anatomy of the retina, whereas Herring’s theory is true of a higher processing level that appeals to opponent mechanisms in neural firing. In a similar way, I have argued that the “trichromatic” theory of elements (I, U, A) is motivated by the anatomy of the articulators, whereas the “opponent theory (I/U and ∀/A) reflect a higher, cognitive, level of organization which is, presumably, likewise guided by neural mechanisms.

References


The Opponent Principle in RcvP


Hulst, H. van der (To appear a). The laryngeal class in RcvP and voice phenomena in Dutch. In: [Festschrift, title presently unknown].


8

Why the Palatal Glide Is Not a Consonant in Japanese
A Dependency-based Analysis

Kuniya Nasukawa

1 Introduction

This chapter challenges the view that the palatal glide \( j \) in Japanese is an onset segment and is the only consonant that may appear in the second position of an initial CC sequence. As we will see, the phonological behavior of \( j \) reveals a strong correlation with the following vowel rather than with the preceding consonant: \( j \) appears if and only if the following vowel is non-front. This dependency relation between \( j \) and a particular class of vowels suggests that the glide forms part of the vowel and not part of the consonant sequence. Furthermore, I argue that an initial \( j \) glide also forms part of the following vowel, making \( jV \) (\( ja, ju, jo \)) a light diphthong rather than a CV sequence.

The discussion also considers what kind of segmental structure is appropriate for representing the palatal glide in Japanese. If the palatal glide forms the initial portion of a light diphthong, then it is reasonable to expect it to be represented as the left-hand part of a contour structure (a branching structure under one timing position). However, in response to several points raised with regard to precedence relations within contour expressions (Lombardi 1990; Schafer 1995; Scobbie 1997; Nasukawa 2005), this chapter argues instead that the sequence \( jV \) is actually the phonetic by-product of a single complex structure which is defined by dependency relations between melodic primes. For example, \( ja \) is the phonetic interpretation of a compound consisting of the melodic primes \(|dlp| \) \(|I|\) and \(|mAss| \)|A|\) (Harris and Lindsey 2000; Harris 2005; Nasukawa Forthcoming), where \(|dlp| \) is structurally dependent on \(|mAss| \) (cf. the mid front vowel \( e \), in which \(|mAss| \) is dependent on \(|dlp| \)). After discussing the validity of this type of representation for
the set of possible glide-vowel sequences in Japanese, the chapter ultimately concludes that some segments such as the Japanese \( j \) can no longer be regarded as formal representational units.

The following analysis introduces a dependency-based model of Element Theory (an element-based feature theory), further developed in Nasukawa (Forthcoming), which regards primes or “elements” as the minimal units of both phonological contrast and phonetic interpretation. This approach views elements as the building blocks of phonological structure, such that combining a particular element with another element is regarded as the foundation of structure-building operations. This differs from intra-segmental structure of the kind employed in models of Feature Geometry, where dependency relations between features are structurally determined and there is no bottom-up construction by combining features (Ewen 1995; Botma, Kula, and Nasukawa 2011).

This chapter is organized as follows. Section 2 describes the status of the segment in different theories with regard to lexical contrast and phonetic interpretation. Then section 3 discusses the standard representation of the palatal glide \( j \) in Japanese and addresses some points of dispute. Section 4 analyzes the representation of the palatal glide \( j \) (and also the (labio-)velar glide \( w \)) in Japanese, taking the view that features are contrastive primes that are independently interpretable. The discussion is concluded in section 5.

## 2 The segment in phonology

### 2.1 The roles of the segment

The segment, which is usually written as a single alphabetic symbol, has always played a central role in phonological theories. Segments are generally thought to lie at the interface between prosody (suprasegmental organization) and melody (segmental architecture): while they often function as the terminal units of prosodic structure, they can also be simultaneously viewed as bundles of melodic features. Phonological studies have employed the segment in different ways as a means of fulfilling one or more of the roles in (1):

(1) The segment is:
   a. a minimal unit of lexical contrast
   b. a minimal unit of phonetic interpretation

These are both indispensable notions in phonological description: the first refers to the most fundamental concept in phonological thinking while the second concerns what kind of phonological information is submitted to the Articulatory-Perceptual systems. These are discussed immediately below.
2.2 Lexical contrast and phonetic interpretation

In classical phonemics (Jones 1950: et passim) the segment is taken to be the minimal unit of phonological contrast (1a) as well as the minimal unit of phonetic interpretation (1b), as shown in (2A). According to this view, features are no longer regarded simply as taxonomic properties of phonemes (cf. Trubetzkoy 1939) and are not units which represent the universal aspects of sounds.

(2) Lexical contrast and phonetic interpretation

<table>
<thead>
<tr>
<th></th>
<th>A Classical phonemics</th>
<th>B SPE, FG</th>
<th>C ET, DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Minimal unit of phonological contrast</td>
<td>the segment</td>
<td>features</td>
<td>features</td>
</tr>
<tr>
<td>b. Minimal unit of phonetic interpretation</td>
<td>the segment</td>
<td>the segment</td>
<td>features</td>
</tr>
</tbody>
</table>

FG = Feature Geometry, ET = Element Theory, DP = Dependency Phonology

In early generative phonology (2B), on the other hand, as exemplified by SPE, the segment (i.e., a full set of distinctive features) still functions as the minimal unit of phonetic interpretation but the feature takes over as the minimal unit of contrast: contrasts are expressed in terms of features, which are understood to be universal properties. In the SPE framework, features must be specified with either a + or – value before being submitted to the Articulatory-Perceptual systems. And even after a given feature acquires a (+/−) value, it cannot be phonetically realized unless it is harnessed to a full set of other (value-specified) features which together determine the phonetic identity of a whole segment.

Yet another view is to be found in frameworks which use monovalent primes ((2C): Anderson and Jones 1974; Anderson and Ewen 1987; Schane 1984; Harris 1999; Nasukawa 2005; Nasukawa and Backley 2008, 2012; Backley 2011; Backley and Nasukawa 2009, 2010), in which the segment does not have any role in expressing contrasts or in shaping phonetic interpretation. Both of the roles in (2) are taken over by features, which are pronounceable individually. For example, in “triangular” theories of melodic representation the vowels i, u, a are the phonetic manifestation of the three melodic features |I|, |U|, and |A|. This idea, which can be traced back to Anderson and Jones (1974), has been developed in various forms in different theories such as Dependency Phonology (Anderson and Ewen 1987: et passim), Government Phonology (Kaye, Lowenstamm, and Vergnaud 1985, 1990; Harris 1990, 1994), Particle Phonology (Schane 1984, 1995, 2005), Radical CV Phonology (van der Hulst 1995, 2005), Strict CVCV theory (Scheer 2004) and Element Theory (Harris and Lindsey...
Why the Palatal Glide Is Not a Consonant in Japanese

1995, 2000; Nasukawa 2005; Harris 2005; Nasukawa and Backley 2008, 2012; Backley and Nasukawa 2009, 2010; Backley 2011). This approach to melodic structure lends itself to monostratal models of phonology, which abandon the underlying-surface distinction and assume that melodic material and also syllable structure are fully specified in lexical representations (Harris 2004; Nasukawa 2010).¹

Taking the (2C) approach in which features are contrastive and independently interpretable, this chapter explores the representation of the palatal glide $j$ in Japanese in section 4. This is preceded by a more general discussion of $j$ which focuses on its phonological behavior and its formal status.

3 The palatal glide $j$ in Japanese

3.1 A general view

The Japanese palatal glide is usually syllabified in an onset and is assumed to be the only consonant that can occur in the second slot of a CC sequence. However, in observing the behavior of $j$ there emerges a strong correlation with the following vowel rather than with the preceding consonant. This dependency of $j$ on the following vowel suggests that $j$ is structurally part of the vowel and not the consonant sequence. On this basis I argue that Japanese $jV$ ($ja$, $ju$, $jo$) constitutes a “light diphthong” rather than a CV sequence. Building on the arguments in Nasukawa (2004, 2005) that (i) a word-final syllabic nasal in Japanese is the phonetic realization of nasality followed by an empty nucleus, and (ii) that Japanese geminates actually have the structure of pseudo-geminates (i.e., two onsets flanking an empty nucleus), the present analysis succeeds in characterizing Japanese as a strict CVCV language in which consonant clusters and consonant-final forms not attested.

3.2 Initial consonant sequences in Japanese

In the literature (Abe 1987: et passim) it is claimed that Japanese allows consonant sequences word/morpheme (syllable)-initially (e.g., $kjookai$ “church”) and word/morpheme-medially (e.g., $kokki$ “national flag”). In both patterns the choice and distribution of segments are restricted. Limiting the discussion to word-initial sequences of the kind shown in (3), the first position allows a wide choice of segments whereas the second position can only be occupied by the palatal glide $j$; that is, $j$ is the only segment permitted to reside in the second slot of a word-initial CC sequence.
In (3), italicized transcriptions are phonemic while the symbols in brackets are phonetic transcriptions which show how coronal obstruents and $h$ merge with $j$ to produce a single palatal “segment” (whereas segments such as labials, velars, and $r$ do not merge with the following $j$).

If we assume the validity of $Cj$ sequences in Japanese, then the structure of the example word $kjokai$ “church” is as follows (cf. Abe 1987: et passim).
In (4), the palatal glide occupies the second position of a syllable-initial branching onset. Although the category “onset” is not formally employed in mora-based models of representation, the structure is essentially the same since the glide occupies the second consonantal Root node of a syllable, as shown below (cf. Labrune 2012: et passim).

Following on from (3), the set of possible initial Cj sequences in Japanese is shown below.

The generalization that emerges from (6) is that a front vowel cannot follow a Cj sequence in Japanese. This is typically accounted for by the co-occurrence restriction in (7), which rules out the sequences *ji and *je.

Interestingly, this is found not only in CjV sequences but also in jV, a single j followed by a vowel.
This suggests that the distributional restriction actually operates between \( j \) and the following vowel, not between a \( Cj \) sequence and the following vowel. Given that the co-occurrence restriction in (7) functions within a constituent – as exemplified by consonant clusters and diphthongs in languages such as English and French – it follows that the glide in a \( CjV \) sequence must be syllabified in the nucleus (\( CjV \)) rather than in the onset (\( Cj-V \)). In fact, any consonant in the Japanese inventory may appear before \( ja, ju, \) and \( jo \), except for \( j, w \) and the placeless syllabic nasal \( N \): the position preceding \( jV \) enjoys the same distributional freedom as a single consonant position which precedes any of the five monophthong vowels \( a, i, u, e, o \).²

(9) Possible initial \( Cj \) in Japanese

<table>
<thead>
<tr>
<th>ka</th>
<th>ki</th>
<th>ku</th>
<th>ke</th>
<th>ko</th>
<th>ga</th>
<th>gi</th>
<th>gu</th>
<th>ge</th>
<th>go</th>
</tr>
</thead>
<tbody>
<tr>
<td>sa</td>
<td>si</td>
<td>su</td>
<td>se</td>
<td>so</td>
<td>za</td>
<td>zi</td>
<td>zu</td>
<td>ze</td>
<td>zo</td>
</tr>
<tr>
<td>ta</td>
<td>ti</td>
<td>tu</td>
<td>te</td>
<td>to</td>
<td>da</td>
<td>di</td>
<td>du</td>
<td>de</td>
<td>do</td>
</tr>
<tr>
<td>ha</td>
<td>hi</td>
<td>hu</td>
<td>he</td>
<td>ho</td>
<td>ba</td>
<td>bi</td>
<td>bu</td>
<td>be</td>
<td>bo</td>
</tr>
<tr>
<td>pa</td>
<td>pi</td>
<td>pu</td>
<td>pe</td>
<td>po</td>
<td>na</td>
<td>ni</td>
<td>nu</td>
<td>ne</td>
<td>no</td>
</tr>
<tr>
<td>ma</td>
<td>mi</td>
<td>mu</td>
<td>me</td>
<td>mo</td>
<td>ra</td>
<td>ri</td>
<td>ru</td>
<td>re</td>
<td>ro</td>
</tr>
</tbody>
</table>

The distributional patterns outlined above raise the question as to what kind of constituent can accommodate \( jV \). Since the sequence contains a vowel, and because \( j \) is a vocoid, presumably \( jV \) as a whole must reside in a nucleus. We are led to conclude, therefore, that \( jV \) is not a \( CV \) sequence after all; it is a light diphthong of the kind found in Chinese and Korean.

This claim finds support in the way these sounds are written in the Japanese syllabary, where \( kj a (きゃ) \) is represented by a combination of \( き (ki) \) and a reduced-size \( ゃ (ja) \) – that is, \( ki \) modified by the addition of \( ja \).

(10) The Japanese syllabary.

\[
\begin{array}{c}
\text{き (ki)} \\
\text{ゃ (ja)} \\
\text{ゅ (ju)} \\
\text{ょ (jo)} \\
\text{きゃ (kja)} \\
\text{きゅ (kju)} \\
\text{きょ (kjo)} \\
\end{array}
\]

In the next section I develop an analysis of the phonological representation of \( j \) which departs from previous treatments. I employ a dependency-oriented model of
Element Theory (an element-based feature theory: Nasukawa Forthcoming) which will shed light on the reason why \( j \) cannot be followed by front vowels in Japanese.

4 Representing the Palatal glide \( j \)

4.1 Previous solutions for \( *ji \) and \( *je \)

It is often claimed that the OCP bans two consecutive tokens of a feature, such as [palatal] or [front], from appearing at the same level of structure. Applying this to the issue at hand, there are at least two possible analyses: in (11a) \( *[\text{pal}][\text{pal}] \) is assumed to operate at the segmental level, while in (11b) the OCP is assumed to apply at the syllable level.

\[
\begin{align*}
(11) \quad \text{a.} & \quad \ast \quad \text{ON} \quad \text{X} \quad \text{OCP} \quad \ast \quad \text{[pal]} \quad \text{[pal]} \quad \text{ji} / e \\
& \quad \ast \quad \text{[pal]} \quad \text{[pal]} \\
\end{align*}
\]

\[
\begin{align*}
\quad & \quad \ast \quad \text{O} \quad \text{N} \quad \text{\sigma} \quad \text{\OCP} \quad \ast \quad \text{[pal]} \quad \text{[pal]} \\
\quad & \quad \text{x1} \quad \text{x2} \quad \text{x3} \\
& \quad \text{C} \quad \text{j} \quad \text{i/e} \\
\quad & \quad \text{[pal]} \quad \text{[pal]} \\
\quad & \quad \text{[pal]} \quad \text{[pal]} \\
\end{align*}
\]

However, it seems that both analyses require some additional explanation to support the proposed representations. In the case of (11a) it is unclear why \( j \) followed by \( i/e \) (i.e., \( *ji, *je \)) is disallowed while \( i/e \) followed by \( j \) (e.g., \( \text{i}j\text{a}ni \) “sarcasm,” \( \text{me}j\text{a}ni \) “eye mucus”) is grammatical. This asymmetry cannot be explained if the OCP (the Obligatory Contour Principle, which prohibits adjacent identical objects: Leben 1973; Goldsmith 1976; McCarthy 1986; Yip 1988) operates only at the segmental level.\(^3\)

Meanwhile, it is questionable whether the \( *[\text{pal}][\text{pal}] \) sequence in (11b) can legitimately be seen as a property of the syllable, as no explanation is offered for why the OCP functions between \( X_3 \) (the head of the domain) and \( X_2 \) (its indirect dependent (via \( X_1 \))): clearly, there is no direct dependency relation between \( X_2 \) and \( X_3 \).\(^4\)

Furthermore, despite the possibility of some phonetic alternation (typically involving a consonant before \( i \) and sometimes before \( u \), as shown in (9): cf. Shibatani 1990: et passim), there are no distributional restrictions between a single onset consonant and a following vowel in Japanese (except for \( jV \) sequences). This implies that the syllable is not a formal constituent in this language (cf. Labrune 2012: Chapter 6).\(^5\)

Ideally we would prefer not to call upon the syllable just for the sake of explaining the phenomenon involving \( *[\text{pal}][\text{pal}] \).

On the other hand, if \( jV \) is analyzed as a light diphthong, it is possible to say that the co-occurrence restriction \( *[\text{pal}][\text{pal}] \) operates at the nuclear level and therefore applies legitimately within a constituent.\(^6\)
It should be noted that this type of contour expression has been called into question by Lombardi (1990), Schafer (1995), Scobbie (1997), Scheer (2003), Nasukawa (2005), Nasukawa and Backley (2008) and others, who raise questions concerning precedence relations between intra-segmental units such as the features present in affricates and prenasalized obstruents (cf. Nasukawa and Backley 2008: 35–36; Nasukawa 2011: 280–283). First, it is difficult to provide an explanation for why affricate contours defy typical edge effects (Lombardi 1990). Second, there is no clear explanation for why the two features in a contour (e.g., [ʤ]) never appear in the reverse order (e.g., *[ʒd]). Third, there is no accounting for the fact that the number of sub-segmental timing slots in an affricate is always two, rather than three, or four, or five. In the absence of a sound reason why contours never contain three or more slots, it seems both arbitrary and accidental that the upper limit should be two.

There is now a growing body of literature addressing these issues, and the view that is emerging is that precedence relations observed in contour segments are not attributable to any sequential ordering of melodic units in a single segmental structure: rather, a contour expression is to be understood simply as the phonetic by-product of a complex structure that is defined by dependency relations between the melodic units contained within it. In the framework of Feature Geometry, for example, Lombardi (1990) claims that a contour expression is a (feature-geometric) structure containing two unordered privative stricture features, [cont] and [stop], each of which resides on a separate autosegmental tier. A dependency relation between these two features determines the phonetic realization of the structurally complex expression.

A similar argument is put forward in Nasukawa (2005) by employing segmental representations based on the version of Element Theory. In this approach (Nasukawa and Backley 2008; Backley and Nasukawa 2009, 2010; Backley 2011), each phonological element (feature) is monovalent and therefore creates privative oppositions; each
is also fully interpretable on its own, and as such, does not require any support from other elements. The elements themselves are listed in (14) with their principal phonetic properties (the label for each element is given in round brackets).

(14) Elements

<table>
<thead>
<tr>
<th>Onset</th>
<th>Nucleus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dental, palatal POA</td>
</tr>
<tr>
<td></td>
<td>labial, velar POA</td>
</tr>
<tr>
<td></td>
<td>uvula, pharyngeal POA</td>
</tr>
<tr>
<td></td>
<td>oral or glottal occlusion, creaky voice (laryngealized vowels)</td>
</tr>
<tr>
<td></td>
<td>aspiration, voicelessness</td>
</tr>
<tr>
<td></td>
<td>nasality, obstruent voicing</td>
</tr>
</tbody>
</table>

Elements are not tied to particular syllabic positions – in principle, any element can appear in any position. However, the same element displays different phonetic properties according to the position where it does appear. For example, in onsets the elements |I|, |U|, and |A| represent the resonance properties which encode differences in place of articulation, while in syllable nuclei they correlate with the vowel categories front, rounded and non-high, respectively. As single elements in a nucleus they are associated with the peripheral vowels [i], [u], and [a], respectively.

(15) a. i  b. u  c. a

| I | U | A |

In most cases, however, segments are represented by compound expressions containing more than one element. For example, a compound comprising the two elements |I| and |A| is phonetically interpreted as a mid front vowel. In the case of Japanese, the compound |I A| manifests itself as the high-mid front vowel e, with |I| and |A| forming a head-dependency relation such that |I| is headed and |A| is its dependent. For example, verb stems must end either in i (e.g., mi “see”) or in e (e.g., ne “sleep”).

(16) Representing e and o in Japanese.

a. e  b. o

| I | U |
| I | A |

In a similar way, the non-front mid vowel o in Japanese is represented as a compound of |U| and |A|, with |U| as the head and |A| as its dependent. Its structure is shown in (16b).

Not only nuclear expressions but also non-nuclear expressions are represented by head-dependent structure (Nasukawa 2005). For example, the difference between prenasalized obstruents (contour segments) and voiced obstruents (non-contour segments) comes down to dependency relations between elements in intra-segmental organization: the element |N| is non-headed in prenasalized stops, as in (17a), but headed in voiced obstruents, as in (17b).
Nasukawa and Backley (2008) also take the view that affrication is entirely a matter of phonetic manifestation and that there is no segment-internal ordering of primes. They conclude that affrication is no more than a performance device designed to improve the perceptibility of complex place cues (e.g., |I| and |A| in (18)), which makes stops with complex resonance more accessible to listeners. This is achieved by enhancing the portion of the speech signal containing aperiodic noise energy, which is relatively rich in place cues.

In this chapter too it is assumed that contour expressions do not exist intra-sententially. In addition, however, it is claimed that the sequence $jV$ is also the phonetic manifestation of a single melodic expression in which an asymmetric (dependency) relation holds between its constituent primes. Taking $ja$ as an example, this must be represented by an expression containing $|I|$ and $|A|$, the same elements that combine in the mid vowel $e$ shown in (16a). Given that $ja$ and $e$ involve the same elements and can therefore only be distinguished by different dependency relations,
I propose that by reversing the head-dependency relation for \(e\) we arrive at the representation for \(ja\).

\[
(19) \quad \text{\(ja\) and \(e\) in Japanese.} \\
\begin{align*}
\text{a. } ja & \quad \text{b. } e = (16a) \\
\begin{array}{c|c|c}
|A| & |I| & |A| \\
\end{array} & \begin{array}{c|c|c}
|I| & |A| & |I| \\
\end{array}
\end{align*}
\]

Here I assume that \(ja\) is the phonetic result of realizing a complex structure which is defined by a particular dependency relation between \(|I|\) and \(|A|\). Note that the structure in (19b) cannot be interpreted as \(aj\) because Japanese has no (heavy) diphthongs. The phonetic difference between (19a) and (19b) is determined on a language-specific basis. In the case of Italian, for example, the structure in (19a) is realized as the low-mid front vowel \(\varepsilon\), whereas in English the same structure is interpreted as a low front vowel \(\text{æ}\).

In the Element Theory and Government Phonology literature (Yoshida 1996: et passim), no conclusion has been reached as to why Japanese employs the structure in (19b) but not that in (19a). Since there is no contrast between \(e\) and \(\varepsilon\) in Japanese, nor between \(e\) and \(\text{æ}\), it is not unnatural to assume that the structure in (19a) is phonetically interpreted as \(ja\). Then by extension, the structures of other \(jV\) light diphthongs are as in (20).

\[
(20) \quad \begin{align*}
\text{a. } ja &= (19a) \\
\text{b. } ju & \\
\begin{array}{c|c|c}
|A| & |I| & |A| \\
\end{array} & \begin{array}{c|c|c}
|U| & |I| & |A| \\
\end{array} & \begin{array}{c|c|c}
|U| & |A| & |I| \\
\end{array}
\end{align*}
\]

Parallel to the \(|A|\)-headed compound of \(|A|\) and \(|I|\) in (20a), the structure in (20b) combines \(|U|\) and \(|I|\) with \(|U|\) as its head. Headed \(|U|\) is reflected in the phonetic realization of the compound, in which a distinct \(u\) vowel (more precisely, \(\text{ɯ}\)) is heard. As for the representation of \(jo\) in (20c), it can be seen that \(|A|\) is a dependent of \(|U|\), but in addition, that \(|A|\) has its own dependent, the element \(|I|\): the most embedded part of the structure (\(|I|\)) is realized as \(j\) while the remaining part of the structure (i.e., an \(|U|\)-headed compound of \(|U|\) and \(|A|\)) corresponds to \(o\), as illustrated in (16b). Japanese is considered to be a five-vowel system, so logically, two further \(jV\) combinations ought to be possible: \(ji\) and \(je\). However, both are ill-formed in Japanese. Employing the same type of structure as above, the ill-formed sequences \(*ji\) and \(*je\) are represented as in (21).
The structure for \( ji \) in (21a) shows two identical tokens of the \( |I| \) element, one of which is dependent on the other. Meanwhile, the structure for \( je \) in (21b) is almost the same as that for \( jo \) in (19c), since they differ only in terms of their ultimate head: (20c) is \( |U| \)-headed while (21b) is \( |I| \)-headed. It is therefore necessary to determine why the structures in (20) are grammatical whereas those in (21) are not.

It appears that, in both of the structures in (21), there operates a co-occurrence restriction on identity avoidance which may be expressed as \( *|I| \rightarrow |I| \).

\[
(22) \quad *|I| \rightarrow |U| \\
\] |I| cannot (in)directly license itself within the same domain.

Unlike the structures in (11), the statement in (22) functions within a constituent domain.9

4.2 The labio-velar glide and vowel

The argument developed above may be extended to the representation \( wV \) (a (labio-)-velar glide \( w \) followed by a vowel) in Japanese. The relevant structures are given in (23).

\[
(23) \quad \begin{align*}
\text{a.} & \quad \text{wa} & \text{b.} & \quad *wi & \text{c.} & \quad *wu & \text{d.} & \quad *we & \text{e.} & \quad *wo \\
|A| & \quad |I| & \quad |U| & \quad |I| & \quad |U| & \quad |I| & \quad |U| & \quad |I| & \quad |U| \\
& \quad |U| & \quad |U| & \quad |U| & \quad |U| & \quad |U| & \quad |U| & \quad |U| & \quad |U| & \quad |U| \\
& \quad |A| & \quad |U| & \quad |U| & \quad |U| & \quad |U| & \quad |U| & \quad |A| & \quad |U| & \quad |U| \\
& \quad |A| & \quad |U| & \quad |A| & \quad |U| & \quad |A| & \quad |U| & \quad |A| & \quad |U| & \quad |A| \\
\end{align*}
\]

Among the structures in (23), only that for \( wa \) is well-formed in Japanese.10 The others may be ruled out using the following statements.

\[
(24) \quad \begin{align*}
\text{a.} & \quad *|U| \rightarrow |U| \\
|U| & \quad \text{cannot (in)directly license itself within the same domain. \( *wu, *wo \)} \\
\text{b.} & \quad *|I| \rightarrow |U| \\
|I| & \quad \text{cannot (in)directly license \( |U| \) within the same domain. \( *wi, *we \)}
\end{align*}
\]
(24a) states that |U| cannot be a dependent when the ultimate head of the expression is also |U|. This excludes *wu and *wo, both of which are |U|-headed but also contain |U| as a dependent. (24b) disallows a dependent |U| in expressions that are headed by |I|.

(25) a. wa  b. *wi  c. *wu  d. *we  e. *wo

The ill-formed *wi in (23b)/(25b) and *we in (23d)/(25d) are also accounted for by (24b): in both cases a headed |I| is prevented from licensing a dependent |U|. Note that the constraints in (24) achieve some degree of parametric variation by allowing each variable to take on different values. Unlike feature-based theories, however, Element Theory succeeds in avoiding over-generation since the range of values (i.e., the number of elements) is relatively small – as few as six in Nasukawa and Backley (2008) and Backley (2011).

In Pöchtrager (2011) it is assumed that phonology parallels syntax to the extent that both share a number of fundamental concepts to do with structure and labeling. And interestingly, Pöchtrager proposes a co-occurrence restriction similar to the one just outlined, in the context of a radical version of Government Phonology that incorporates structural asymmetry between elements. His approach defines structural asymmetry in terms of the established syntactic notion of binding: a structurally-embedded element must be bound by another element. According to Pöchtrager (2011), the co-occurrence restrictions in (22) and (24) are captured by binding relations between elements. His model differs from the proposed representations in (20), (21), and (23) in that the glides j, w and their following vowels are assumed to be in the spec x1 and the head xN respectively, as illustrated (N = nucleus, NP = max. projection, x1 = spec., X2 = comp.).

(26) a. *ji  b. ju  c. *wi
Without going into detail about the differences between Pöchtrager’s structures in (26) and the representations proposed in this chapter, it is interesting to note that Pöchtrager (2011) employs the same co-occurrence restrictions as those introduced in (22) and (24). In Pöchtrager’s terms, (22) and (24a) are accounted for by “no self-binding,” which states that an element cannot bind a token of itself. On the other hand, his version of (24b) is expressed by a constraint stating that |U| cannot bind |I|. He illustrates the validity of these restrictions by referring to other phenomena from different languages including English, Putonghua, and Turkish.

So although the structure used by Pöchtrager differs from the one proposed in this chapter, the two approaches are similar in that both employ the same type of co-occurrence restriction controlling asymmetric relations between elements. On the other hand, what distinguishes the current approach to co-occurrence restrictions from those found in feature-based theories is its position regarding over-generation, as mentioned above. The model developed here generates only a limited number of co-occurrence restrictions since the number of available elements is relatively small, whereas feature-based theories have the potential to describe a much larger number of possible restrictions since the feature set itself is relatively large. They must then explain why most of these co-occurrence restrictions are never actually attested – for example, by making appeal to other grammatical devices such as markedness theory.

By employing the proposed dependency-based structure for jV and wV ((20), (21) and (23)), it becomes possible to analyze Japanese as a strict CV language. This outcome is supported by arguments made in the Government Phonology literature (e.g., Nasukawa 2004, 2005; Takahashi 2004) that (i) Japanese (full and partial) geminates actually have the structure of pseudo-geminates (i.e., two onsets flanking an empty nucleus); (ii) a syllable-final moraic nasal n in Japanese is the phonetic realization of a CV structure in which the C slot contains only nasality and the V slot is melodically empty; and (iii) V₁V₁ and V₁V₂ are to be viewed as sequences of nuclei and not as long vowels or diphthongs. In short, consonant sequences do not occur in Japanese. In this chapter, I have presented my analysis using element-based representations, as these allow for a clear and elegant description. However, the same result is achievable within standard feature-based approaches too, thus reinforcing the point that my conclusion primarily concerns the structure of Japanese and not the workings of Element Theory itself. An additional outcome is that we must revise our understanding of the Japanese vowel system: rather than assuming a simple system of five vowels, we must recognize a total of nine expressions that can occupy a nucleus – five monophthongs (a, i, u, e, o) and four light diphthongs (ja, ju, jo, wa).

5 Conclusion

This discussion has been based on a version of Element Theory which employs a small set of phonological primes called elements. In this approach, precedence relations between the elements in a contour expression are not specified. Instead, the
dependency relations holding between elements determine how a structure is phonetically realized. Thus, the linear ordering of segmental material is to be viewed as no more than the phonetic by-product of a particular phonological representation.

In the context of the full-dependency model of Element Theory, this chapter concludes that Japanese has a vowel system which is larger than the one generally assumed. In addition to the standard five vowels (a, i, u, e, o) it is necessary to include the light diphthongs ja, ju, jo, wa, which have traditionally been syllabified as CV sequences rather than complex nuclei. This analysis aligns with current thinking regarding other East-Asian languages such as Chinese and Korean, where the same patterns are treated as light diphthongs rather than as glide-vowel sequences.

On a final note, in present-day Japanese (excluding elderly speakers) the katakana spellings of many foreign proper names suggest that the sequences je, wi, we, and wo are (at least phonetically) possible. In response to this, we may assume that a lexical level in Japanese often referred to as the “foreign stratum” employs a set of co-occurrence restrictions different from those proposed above. In the foreign lexical stratum, only the restrictions in (22) and (24a) are apparently active. These can be subsumed under a restricted version of Pöchtrager’s constraint “no self-binding” (an element cannot bind a token of itself): specifically, it must be recast as “no direct self-binding,” thereby filtering out the ungrammatical patterns *ji and *wi. Meanwhile, the actual pronunciation of the sequences je, wi, we, wo is uncertain and inconsistent (Vance 2008). This implies that the constraint “no direct self-binding” has not yet established itself in native speaker’s mental grammar. This is clearly an area for future research.

Acknowledgments

The ideas proposed in this chapter were first presented at the CUNY Conference on the Segment held at CUNY Graduate Center in January 2012. My thanks go to the conference participants for their constructive comments, and to Phillip Backley, Charles Cairns, and Eric Raimy for discussion and corrections of an earlier draft. This research was partially supported by a Grant-in-Aid for Scientific Research (B) from Japan Society for the Promotion of Science, Grant No. 22320090.

Notes

1 Even in the (2C) approach, however, the notion “segment” is not discarded entirely: features are attached to segment-size units (known as skeletal positions, CV units, or root nodes) which have a key role in representing precedence relations at the segmental level. In formal terms, then, segments still have work to do. For a detailed discussion, see Nasukawa (2011).

2 The restriction on j may be analyzed along the lines of Cairns and Feinstein (1982), where the only phoneme that can occur in the “adjunct” position of the Onset is j and where there is a ban on the “adjunct” position before front vowels.

3 For discussion see Yip (1988, 1998), van Riemsdijk (2008) and Nasukawa and Backley (Forthcoming).
The problem of *[pal][pal] only arises in models of phonology in which precedence relations have no formal significance. Unless one denies that phonology cannot express the idea of “immediately before” or “immediately after” then this constraint is not problematic.

In the Government Phonology framework (Kaye, Lowenstamm, and Vergnaud 1990; Harris 1994), the syllable is not recognized as a formal constituent in any language.

Japanese displays other distributional restrictions which operate within the domain of a particular constituent: e.g., *[voi][voi] at the morpheme level and minimal length at the foot (foot binarity).

The notion of asymmetric relations is captured not only by the term dependency, but also by terms such as head-complement, licensing, and weak/strong (for references, see Nasukawa 1995).

Apparent diphthongs such as *ai are best analyzed as sequences of two nuclei. Unlike English diphthongs, vowel sequences in Japanese freely combine any two vowels in either order.

For a discussion of the nature of this type of negative constraint, see Nasukawa and Backley (2010, Forthcoming). It is beyond the scope of this chapter to examine in detail the nature and/or validity of this type of constraint.

Eric Raimy has pointed out that the existence of CJa words predicts the possibility of Cwa words in Japanese too. It has been claimed (Frellesvig and Whitman 2004) that at an earlier stage in the history of the language there existed distributionally-restricted sequences such as kwa and kwo, which have now disappeared. In Element Theory terms, this indicates a tendency to avoid the complement [U] (cf. other elements, and also headed [U] in wu and o), its presence apparently affecting the ability of elements to occupy spec positions. This may be related to the phonological weakness of wu in Japanese (Nasukawa 2010, Forthcoming).

References


Determining Cross-Linguistic Phonological Similarity Between Segments

The Primacy of Abstract Aspects of Similarity

Charles B. Chang

1 Introduction

The notion of phonological similarity – that is, similarity between two sound structures – is central to much of the literature on spoken language. Phonological similarity is invoked to explain a variety of systematic patterns in word recall (e.g., Copeland and Radvansky 2001; Fournet et al. 2003), lexical and conceptual development (e.g., Sloutsky and Fisher 2012), language games (e.g., Zwicky and Zwicky 1986), first-language (L1) and second-language (L2) perception (e.g., Johnson 2003; Best and Tyler 2007), L1 and L2 production (e.g., Major 1987; Page et al. 2007), loanword phonology (e.g., Kang 2003, 2008), and cross-linguistic interaction in bilingualism (e.g., Flege 1995; Laeufer 1996).

Several different kinds of “phonological similarity” are referred to in the literature, however, and these various types of similarity have diverse consequences for grammar and learning (for a recent overview, see Gallagher and Graff 2012). For example, some studies examine the effects of phonological similarity between lexical items – operationalized as “neighborhood density” – on speech perception and production (e.g., Luce and Pisoni 1998; Vitevitch 2002; Munson and Solomon 2004; Gahl et al. 2012), while other studies consider the similarity between the various potential forms of a lexical item in explaining distributional regularities such as phonotactic restrictions and environments for alternation and neutralization (e.g., Pierrehumbert 1993; Flemming 2004; Frisch et al. 2004; Steriade 2009; Gallagher 2012). Phonological similarity between individual sounds or natural classes of sounds has been measured perceptually via perceptual confusions or explicit mappings with goodness-of-fit ratings (e.g., Miller and Nicely 1955; Strange 1999; Best et al. 2003; Chang 2009b). Computational work, on the other hand, draws on a feature-based
type of phonological similarity to align segmental sequences, whether for the purposes of analyzing cognate relationships or developmental speech patterns (e.g., Covington 1996; Kondrak 2003; Kessler 2005). Importantly, a similarity metric that provides a good model of behavior in one case may make poor predictions in another. As noted by Gallagher and Graff (2012), perception and production data do not necessarily converge on the same similarity relations, nor do phonetic and phonological data (Mielke 2012).

The mismatch between “phonetic” kinds of similarity and “phonological” kinds of similarity is at the heart of a disparity that is commonly seen between segmental similarity relations within one language and those between two languages. In this chapter, I describe this mismatch in more detail and argue that conflicts between different types of similarity are so often resolved in the same way (namely, in favor of “phonological” kinds of similarity) because high-level information is weighed more heavily than low-level information. Note that the segment is fundamental to this argument because there is no clear way to implement the phonemic-level interactions described in this chapter without positing an abstract, segment-sized category such as the phoneme. In a sense, then, the cross-linguistic phenomena reviewed here can be considered evidence for the existence of segments as discrete phonological units, as well as for the very distinction between phonetics and phonology (in particular, their hierarchical relationship).1

The chapter is organized as follows. In section 2, I decompose the construct of phonological similarity into subtypes of similarity and review the problem of conflict between different types of similarity observed in the cross-linguistic speech literature. In section 3, I present an array of findings from cross-linguistic research showing a preference for relating segments and natural classes to each other on an abstract level. In section 4, I discuss the implications of such abstract knowledge for studies of cross-linguistic phonetics and phonology, and in section 5, I provide concluding remarks.

2 Components of phonological similarity and their interaction

2.1 Levels of similarity

The construct of phonological similarity can be decomposed into at least2 three subtypes of similarity: objective acoustic similarity, language-specific allophonic similarity, and cross-linguistic phonemic similarity. These metrics of similarity have analogues in other models of phonological similarity that distinguish between various factors influencing overall similarity (e.g., Austin 1957; Ladefoged 1969; Flege 1996; Bohn 2002). Let us consider each type of similarity in turn.

Acoustic similarity refers to the raw (i.e., non-language-specific) distance between sounds in terms of acoustic dimensions such as frequency, duration, and amplitude. At a basic auditory level, listeners tend to perceive sounds that are
relatively close acoustically (e.g., [f] and [θ]; [i] and [ɪ]) as more similar than sounds that are relatively distant acoustically (e.g., [s] and [θ]; [i] and [a]). For example, in a speeded discrimination task (thought to reflect non-linguistic perception of auditory contrast), native Dutch speakers and native English speakers take a comparably longer amount of time to discriminate the acoustically similar [f] and [θ] than the acoustically dissimilar [s] and [θ], even though these pairs of sounds are contrastive in English only (Johnson and Babel, 2010). This kind of result is consistent with the view that there is an acoustic/auditory basis for perceived similarity that transcends linguistic knowledge. Nevertheless, differences in language background are likely to result in divergent perceptual patterns in linguistic tasks due to effects of allophonic similarity.

Allophonic similarity is based on within-language comparisons between sounds at the level of contextually defined allophones, which are specific to a particular language. A pair of sounds is allophonically similar to the extent that they can be related to each other within a language – by virtue of the fact that they do not contrast and/or the fact that they alternate with each other in a productive pattern (Johnson and Babel 2010). Consequently, a pair of sounds can be perceived differently by listeners of different language backgrounds if the two sounds exist in an allophonic relationship in one language, but not the other. For example, English speakers (for whom [d] contrasts with [ð] and alternates with [ɾ]) perceive [d] as more similar to [ɾ] than to [ð]; in contrast, Spanish speakers (for whom [d] contrasts with [r] and alternates with [ð]) perceive [d] as more similar to [ð] than to [r] (Boomershine et al. 2008). Similar patterns are found when a sound is absent from one language, but present in another. For instance, Dutch speakers (for whom [θ] does not occur as a phoneme) rate [θ] as more similar to [s] and [ʃ] than do English speakers (for whom all three fricatives are phonemes) (Johnson and Babel 2010).

While acoustic similarity is not specific to any language and allophonic similarity is specific to one language, phonemic similarity is related to sounds in two languages. Therefore, phonemic similarity is inherently cross-linguistic. Phonemic similarity is also abstract, because it is based on cross-language comparisons between sounds at the level of context-free phonemes, which may be viewed in terms of feature bundles. A pair of sounds that are acoustically and/or allophonically dissimilar may nonetheless be phonemically similar due to at least two factors: (1) similar positions in the respective phonemic inventories (when considering the contrastive feature oppositions – or, more broadly, the “relative phonetics” – of the sounds in relation to other sounds in the inventory), and (2) similar distributional facts.

To take an example, American English and Mandarin Chinese both contain a vowel standardly transcribed as /u/ in their respective inventories, but the languages differ substantially in the quality of their /u/. English /u/ is acoustically far from Mandarin /u/ and much closer to the Mandarin front rounded vowel /y/ (Chang et al. 2011). Nevertheless, these two /u/ vowels can be identified as the “same” phoneme because they each occupy a similar place within the relevant vowel inventory – namely, that of a high back rounded vowel (i.e., [-consonantal, +syllabic, +high, +back, +round]). Even though English /u/ is relatively front and unrounded
in comparison to Mandarin /u/, it is still the vowel that is the most high/back/ rounded in the English inventory and, therefore, the most parallel to Mandarin /u/ in terms of vowel features. In addition to parallel inventory niches, English /u/ and Mandarin /u/ show similar distributional restrictions with the back rounded approximant /w/: neither can occur with /w/ in a stop-approximant onset cluster (i.e., *[pwu], *[twu], *[kwu], etc.). These similar co-occurrence restrictions suggest that, even though English /u/ is acoustically quite far from Mandarin /u/, they both pattern like back rounded vowels. In this way, English /u/ and Mandarin /u/ are phonemically similar despite their disparate phonetic realizations.

In summary, phonological similarity between segments can be said to exist at multiple levels: acoustic, allophonic, and phonemic. Acoustic similarity and allophonic similarity are relevant both within and between languages; however, phonemic similarity, since it involves the comparison of two phonological systems, is relevant only for cross-linguistic comparisons. As such, the perceived similarity between two segments within a language has typically been discussed in acoustic and/or auditory terms. In the next section, we review one influential attempt to encode this kind of perceptual similarity in the grammar and show how its predictions break down if extended to cross-linguistic comparisons.

2.2 Perceptual similarity in a (monolingual) grammar

Given how often linguistic phenomena are explained in terms of phonological similarity, it is reasonable to think that knowledge of similarity constitutes part of linguistic knowledge, and Steriade (2009) attempted to represent this knowledge in a language-universal “P-map,” a set of ranked constraints regarding the relation of relatively similar vs. dissimilar forms. These constraints aim to maximize perceptual similarity between input and output forms, such that input-output correspondences between relatively dissimilar forms are penalized more heavily than those between relatively similar forms. For instance, in the case of final voiced stops, a typologically dispreferred structure, there might be two constraints – one penalizing a correspondence between a syllable-final voiced stop and a voiced stop-initial syllable containing an epenthetic vowel (*D[V]o–DV[V]o) and another penalizing a correspondence between a syllable-final voiced stop and a syllable-final voiceless stop (*D[V]o–T[V]o). Because a syllable-final voiced stop is arguably less similar to a new syllable than to a syllable-final voiceless stop (as reflected in similarity judgment data; Kawahara and Garvey 2010), the first constraint *D[V]o–DV[V]o is ranked above the second constraint *D[V]o–T[V]o (*D[V]o–DV[V]o >> *D[V]o–T[V]o), such that, absent the influence of intervening constraints, syllable-final voiced stops are predicted to alternate with syllable-final voiceless stops, not with syllables containing an epenthetic vowel. In fact, this is what is found across a range of languages (e.g., Germanic and Slavic languages): final voiced stops are repaired via devoicing rather than epenthesis. This pattern is thus consistent with the basic prediction of the P-map – namely, that output patterns follow perceptual similarity relations between an input and its possible outputs.
But what about mapping between an L2 input and an L1 output? In a recent study, Shaw and Davidson (2011) showed that the tight link between perceptual similarity and ultimate output assumed by the P-map does not hold for cross-linguistic mapping. Controlling for a variety of factors, they observed that unfaithful production of novel (L2) input clusters cannot be said to follow from perceptual similarity, as fricative-stop clusters were produced with epenthetic forms (inserting a vowel in the middle of the cluster), despite being judged most similar to prothetic forms (inserting a vowel before the cluster). Some explanations offered for this unexpected disparity between production and perception were maximizing the perceptual recoverability of segments, as well as maintaining uniformity in repair strategy (given that stop-stop clusters were also produced with epenthetic forms). Still, it remains unclear why the P-map, which seems to do a good job accounting for within-language alternation, fails in this kind of cross-linguistic situation, since it is supposed to represent universal perceptual knowledge.

Here I consider the possibility that the P-map fails in cross-linguistic circumstances because, as a model of similarity based on within-language relations, it does not incorporate the influence of phonemic similarity between languages. The ‘P’ in the P-map stands for “perceptual,” which reflects the fact that it encodes similarity relations based on (acoustic) perceptual similarity. However, as discussed above, acoustic similarity and allophonic similarity are not the only types of similarity that exist between L1 and L2 segments. Cross-linguistic phonological similarity may also be influenced by phonemic parallelisms, leaving the P-map ill-equipped to fully model cross-linguistic similarity relations. Thus, in the next section, we examine the hypothesis that phonemic similarity – in particular, its interaction with acoustic and allophonic similarity – plays a primary role in determining cross-linguistic similarity relations that depart from acoustic and allophonic comparisons (and, therefore, from predictions of the P-map).

2.3 Conflict and interaction between levels of similarity

The idea that phonemic similarity can result in cross-linguistic similarity relations differing from within-language similarity relations is based on two assumptions. The first assumption is that phonemic similarity sometimes differs from acoustic and allophonic similarity; as is discussed below in section 3, there is ample evidence that this situation actually arises. The second assumption is that at the phonetics-phonology interface there is a hierarchy between higher-level information (e.g., phonemic correspondence) and lower-level information (e.g., acoustic properties, allophonic alternations), with higher-level information taking precedence in cases of conflict.

In discussing this latter assumption, I proceed from an implication made by Flege (1996) in his definition of how to determine when a novel L2 sound is “new” vs. “similar” to an L1 sound. Flege observed that a useful heuristic in determining L1-L2 similarity is the “phonetic symbol test,” one of Bohn’s (2002) so-called
“armchair methods”: if an L1 sound and an L2 sound are transcribed with the same symbol in the International Phonetic Alphabet, this implies that the L2 sound is not “new,” but rather “similar” or “identical” to the L1 sound. Given that transcription conventions for a given language are often based on phonemic considerations (e.g., the contrastive status of certain phonetic details), such a phonetic symbol test will often resemble a cross-linguistic phonemic analysis. Flege noted that the phonetic symbol test was not absolute, however, and that its results should be supplemented with acoustic and perceptual data in making predictions about the relation of L1 and L2 sounds. The shortcomings of this method were also pointed out by Bohn (2002), who noted that perceptual measures provide the most stable assessments of phonological similarity. Neither Flege nor Bohn specified how different types of similarity should be resolved when they make conflicting predictions.

The hypothesis examined here is that different types of cross-linguistic phonological similarity are resolved by L2 users in favor of higher-level similarity. In other words, the manner in which L2 users relate L2 segments to L1 segments is predicted to be based predominantly upon abstract, between-system comparisons at the phonemic level, not within-system comparisons at the allophonic level or system-external comparisons at a psychoacoustic level. As outlined above, such a cross-linguistic phonemic level of analysis probably considers at least a segment’s position within relevant featural dimensions as well as distributional information. In this regard, it is important to point out that the term “L2 users” is meant to refer to individuals who would have access to this kind of information – that is, L2 users with phonemic knowledge of the L2, not naïve listeners exposed to the L2 for the first time. Relatively experienced L2 users are expected to show L1-L2 mappings that follow phonemic similarity over acoustic and allophonic similarity because of a tendency for high-level information to override low-level information, consistent with many other “top-down” effects in speech processing (e.g., Warren 1970; Jongman et al. 2003; Davis and Johnsrude 2007). In the following section, these predictions are shown to be borne out in a wide range of cross-linguistic research.

3 Phonemic similarity in cross-linguistic research

3.1 Phonemic correspondence in second-language perception

Because languages differ in terms of phonemic inventory and patterns of allophonic alternation, both phonemic similarity and allophonic similarity are language-specific kinds of similarity, as explained in section 2.1. It should thus come as no surprise that L2 perception of a given phonological structure has been observed to vary across L1 backgrounds. The language-specific nature of cross-language mapping is often attributed to the existence of different phonological constraints, or different rankings of constraints, across languages (e.g., Broselow et al. 1998), but some part of this language specificity is likely due to basic cross-linguistic differences in the perception of unfamiliar phonological structures.
The Perceptual Assimilation Model (PAM; Best 1994, 1995) – an articulatory framework for understanding non-native speech perception – has played a particularly influential role in the analysis of cross-linguistic differences in non-native perception, attributing them to the various ways in which non-native sounds may align with the gestural constellations of native phonological categories. The core insight of the PAM lies in relating disparities in perception of foreign contrasts to disparities in native phonological knowledge gained through linguistic experience. Difficulty in discriminating a foreign contrast arises when the structure of a listener’s native phonology interferes, causing the foreign sounds to be perceptually assimilated to a single native category. For example, when clicks – language sounds that are relatively distinct acoustically – are played to listeners with no native click categories, the clicks are discriminated well; in the case of click-language speakers, however, non-native clicks are discriminated poorly, due to convergent perceptual assimilation to native click categories (Best et al. 2003). Perceptual assimilation to L1 structures results in cross-linguistic differences not only in the discriminability of non-native segments, but also in the perception of non-native phonotactics, such as initial and medial consonant clusters (Dupoux et al. 1999; Hallé and Best 2007). These findings demonstrate that L1 phonological structure exerts a profound influence on cross-linguistic speech perception, biasing listeners of different language backgrounds toward interpreting the same L2 input in disparate ways (see also, e.g., Weinreich 1953; Flege 1987, 1995).

However, even though the specific percept of a given L2 segment may usually differ across L1 backgrounds, it is still reasonable to predict that the L2 segment will be perceptually assimilated to the L1 segment that is the closest phonetically (acoustically and/or articulatorily), whatever that may be. Consistent with this prediction, the literature on L2 speech perception includes many findings of close correspondence between acoustics, or phonetic realization more generally, and perceptual performance (see, e.g., Bohn and Best 2012 for recent findings on cross-linguistic discrimination of approximants). Nevertheless, some studies evince a dissociation between acoustic similarity and perceptual similarity. Perception of non-native vowels in particular has been repeatedly shown to bypass the acoustically closest L1 vowels in favor of the phonemically closest L1 vowels. For example, native speakers of Canadian English judge German /u/ to be more similar than German /y/ to English /u/, despite the fact that German /y/ is the acoustically closer vowel to English /u/ (Polka and Bohn 1996). Similarly, native speakers of American English judge front rounded vowels in both French and German to be more similar to English back rounded vowels, even though they are acoustically closer to the English front unrounded vowels in three-dimensional ($F_1 \times F_2 \times F_3$) vowel space (Strange et al. 2004). These findings demonstrate that cross-linguistic perceptual similarity does not follow straightforwardly from traditional measures of acoustic similarity; rather, listeners may perceive an L2 segment as most similar to the phonemically closest L1 segment, even when this is not the acoustically closest one.6

The pattern of perceiving the phonemically closest L1 segment as most similar to an L2 segment is consistent with the idea of perceptual assimilation at the
phonological (abstract) level, as described in a version of the PAM for L2 learners, the PAM-L2 (Best and Tyler 2007). Although the PAM-L2 does not specify how the phonological level interacts with the phonetic level and the gestural level, it does state that this interaction is likely to change over time as L2 learners gain more knowledge of the L2, suggesting that phonemic information may come to play a larger role as more of it becomes available over the course of L2 learning. Indeed, when L2 learners have a modicum of L2 phonemic knowledge, they seem to prioritize phonemic correspondence over acoustic proximity in the calculation of overall cross-linguistic phonological similarity, as is evident in much of the work on L2 production.

3.2 Phonemic correspondence in second-language production

Like studies of L2 perception, studies of L2 production suggest that phonemic similarity plays a large role in relating L2 forms to L1 forms. In the Speech Learning Model (SLM; Flege 1995, 1996), a model of L2 perception and production that assumes an allophonic level of analysis, phonemic similarity is not discussed as such; however, this corresponds closely to what is measured in the phonetic symbol test (see section 2.3): phonemically similar sounds tend to be transcribed with the same symbol. Together with acoustic and perceptual similarity, phonemic similarity helps predict whether novel L2 sounds will be classified by L2 learners as “new,” “similar,” or “identical” with respect to familiar L1 sounds. Classification of an L2 sound as “identical” to an L1 category does not negatively impact its production, as any differences between the two are negligible, whereas classification of an L2 sound as “similar” to an L1 category negatively affects its production because of a significant phonetic disparity between the L1 and L2 sounds. Analogizing an L2 sound to a similar, but non-identical, L1 sound results in their perceptual linkage, which allows the disparate properties of the L1 sound to influence the production of the L2 sound (and vice versa). A “new” L2 sound, by contrast, has no clear L1 counterpart and is thus not analogized to an L1 sound, which allows it to be produced accurately, free from L1 influence, once sufficient experience in the L2 has been acquired.

Although the SLM, like the PAM(-L2), does not address the interaction among the various kinds of similarity that influence overall cross-linguistic similarity, the L2 speech literature implies that, for L2 learners, phonemic similarity takes precedence over acoustic phonetic similarity. Despite the tendency for different metrics of similarity to converge toward the same conclusions, it is not uncommon for phonemic similarity and acoustic similarity to be at odds with each other (see, e.g., Hammarberg 1996), as shown by the frequent disagreement between acoustic comparisons and transcription conventions, which for a given language are based partly on phonemic considerations. Since the SLM relates L1 and L2 sounds at a position-sensitive allophonic level, it seems to predict that L2 learners will resolve such conflicts in favor of acoustic phonetic comparisons; however, this is not how the “new” vs. “similar” distinction is applied throughout the literature, including Flege and Hillenbrand’s (1984) study of American English speakers’ production of French /u/ and French /y/
(in the words *tous* /tu/ “all” and *tu* /ty/ “you”). For L1 English speakers, L2 French /u/ is analyzed as a “similar” vowel with an L1 counterpart in English /u/, while L2 French /y/ is analyzed as a “new” vowel with no L1 counterpart. This classification contrasts with the acoustic facts, which show that in the given alveolar context, English /u/ is actually closer to French /y/ than to French /u/ (Strange *et al.* 2007). Nevertheless, the production of L2 French vowels by L1 English speakers is consistent with the phonemic pairing of vowels: French /y/ is produced accurately, while French /u/ shows influence from English /u/ – a result that has been replicated for L1 English learners of Mandarin (Chang *et al.* 2011).

Natural classes of consonants also show this bias toward phonemic correspondence in L2 production. For example, English-Portuguese bilinguals – influenced by the long-lag voice onset time (VOT) of L1 English voiceless stops – produce the L2 Portuguese voiceless unaspirated stops with relatively long VOT, not with the acoustically more similar short VOT characteristic of L1 English voiced stops (Major 1992). English-French bilinguals show the same effect with L2 French /t/, which they produce with VOT that is too long, under influence from long-lag L1 English /t/ (Flege 1987). In both cases, inauthentic, or accented, VOT production follows from a perceptual linkage of the L2 voiceless stops with the phonemically corresponding L1 voiceless stops, not with the L1 voiced stops (which, with their short-lag VOT, are arguably more similar at an acoustic phonetic level). The pattern that emerges is that L2 users seem to favor linking L1 and L2 sounds on the basis of phonemic correspondence rather than strictly acoustic proximity. This same pattern is found in cross-linguistic influence of the L2 on the L1.

### 3.3 Phonemic correspondence in phonetic drift

According to the SLM, the same mechanism of cross-linguistic perceptual linkage between “similar” L1 and L2 sounds is responsible for both L1-to-L2 interference and L2-to-L1 interference, although the theoretical distinction between “similar” and “new” L2 sounds has rarely been discussed in concrete acoustic terms (cf. Strange 1999 for a typology in terms of mappings and goodness-of-fit ratings). Chang (2010) suggested that, along with “identical” L2 sounds, “similar” and “new” L2 sounds form a continuum of cross-linguistic similarity to the L1, with the boundary between “identical” and “similar” generally corresponding to a cross-linguistic disparity of one just-noticeable difference (JND) in the relevant acoustic dimension. This conceptualization of the L2 in relation to the L1 is found to correlate well with patterns of L1 “phonetic drift” – subtle shifts in L1 speech production following from L2 experience in adulthood (e.g., Flege 1987; Major 1992; Sancier and Fowler 1997).

In a study of L1 English novice learners of Korean, Chang (2010, 2012a) showed that phonetic drift occurs after even brief L2 experience, at multiple levels of phonological structure, and in accordance with initial cross-linguistic distances between the L1 and the L2. Crucially, L1 structures drift toward L2 structures only when
there is an appropriate amount of cross-linguistic distance between them. To trigger phonetic drift, an L2 structure must be similar enough to an L1 structure to qualify as “similar” rather than “new,” yet different enough not to be perceived as “identical” (i.e., at least one JND away from the L1 in the relevant dimension). Thus, for L1 English learners of Korean, the VOT of L1 English voiceless stops drifted toward the VOT of the perceptually similar L2 Korean aspirated stops (which differs by more than the JND for VOT), whereas the VOT of L1 English voiced stops did not drift toward the VOT of the perceptually similar L2 Korean fortis stops (which differs by less than the JND for VOT).

Although the L1-L2 perceptual linkages proposed in Chang (2010, 2012a) are justified on an acoustic basis, these results do not provide clear evidence of a phonemic basis to relations between L1 and L2 sounds, since the relevant natural classes of English and Korean stops differ phonologically in a number of ways and the participants – novice learners – may not have had sufficient phonological knowledge of the L2 to draw phonemic comparisons in any case. Studies of phonetic drift in advanced L2 learners, however, are consistent in showing a preference for linking L1 and L2 sounds that correspond phonemically. For example, French-English bilinguals produce their short-lag L1 French /t/ with overly long VOT, under influence from long-lag L2 English /t/ (Flege 1987). From an acoustic perspective, it would be more favorable for them to link their L1 French /t/ instead to the similarly short-lag L2 English /d/, in which case they would not be expected to manifest much phonetic drift at all; however, the observed pattern of drift instead evinces a perceptual linkage between the phonemically corresponding voiceless stops. Similarly, Portuguese-English bilinguals produce L1 Portuguese voiceless stops with VOT that is influenced by the longer VOT of L2 English voiceless stops, not by the similar short-lag VOT of L2 English voiced stops (Sancier and Fowler 1997). These production data thus provide additional evidence of respect for phonemic correspondence between L1 and L2 stop types, a phenomenon that is also well documented in the literature on loanwords.

3.4 Phonemic correspondence in loanword phonology

The importance of phonemic similarity in determining cross-linguistic mappings has been amply demonstrated in studies on loanword adaptation, as reviewed by Kang (2008, 2011). This literature has shown that when foreign words are phonologically incorporated into a language, acoustic perceptual similarity interacts with phonemic similarity in complex ways. While many modifications to borrowed L2 forms mirror patterns of perceptual “foreign accent” (Peperkamp and Dupoux 2003), other changes do not seem to maximize perceptual similarity (e.g., Shinohara et al. 2011) and map an L2 form to the phonemically most similar L1 form instead of the perceptually most similar one. Similarity between L1 and L2 sounds at the phonemic level is able to play a role in loanword adaptation because the primary agents of adaptation are typically fluent bilinguals acquainted with the phonological
Determining Cross-Linguistic Phonological Similarity Between Segments

structure of both the L1 and the L2, not monolingual L1 speakers (Paradis and LaCharité 1997, 2008). The outcome of loanword adaptation, however, is not fully determined by phonemics, as borrowed forms frequently evince a more fine-grained analysis of the L2 signal. Final voiced stops in English, for instance, are adapted into Korean variably as unreleased voiceless stops or lenis stops with a following epenthetic vowel, depending on the quality of the preceding vowel and the place of articulation of the stop (Kang 2003). This sort of variability reveals a nuanced sensitivity to the phonetics of the lending L2 that would be lost in a strictly phonemic analysis, suggesting that the phonemic representation of a borrowed word is enriched by phonetic detail (Kang 2008), which influences the outcome of adaptation at the same time as phonemic information (Chang 2012b).

Nevertheless, L2-to-L1 mapping in loanword adaptation often evinces a respect for source (L2) phonemics, and many relevant cases of following phonemic correspondence over acoustic proximity are reported in detail by LaCharité and Paradis (2005). Cross-linguistic mapping of vowels, for example, tends to occur on a phonemic basis in loanwords. In the case of English borrowings in Quebec French, the English high lax vowels /ɪ, ʊ/ are acoustically closest to the French mid vowels /ɛ, ɔ/, but these English vowels are consistently mapped to the French high vowels /i, u/, not to the French mid vowels /ɛ, ɔ/ (or /e, o/). In the case of English borrowings in Japanese, the English rhotic /r/ (realized as an alveolar approximant [ɹ]) is mapped onto the Japanese rhotic /r/ (realized as a postalveolar flap [ɽ]), not onto the Japanese approximant /w/, even though /w/ is perceptually more similar. Finally, in the case of English borrowings in Mexican Spanish, the English voiced stops, despite being realized as voiceless word-initially, are nearly always adapted as the strongly prevoiced Spanish voiced stops, not as the voiceless stops (which are more similar in terms of VOT). Similar findings are reported by Chang (2009a, 2012b) for English borrowings in Burmese. Even though English voiceless stops are typically realized as aspirated word-initially, they are nearly always borrowed into Burmese with the Burmese voiceless unaspirated stop series, not with the voiceless aspirated series. These data show overall a respect for source phonemic distinctions, which results in cross-linguistic mapping to phonemically parallel sounds even when these sounds are not the most similar acoustically.

4 Discussion

In this chapter, I have endeavored to make three points: (1) “phonological similarity” is a complex construct consisting of, and influenced by, multiple types of similarity; (2) levels of similarity are hierarchically organized, with high-level similarity ranked above low-level similarity; and (3) the influence of phonemic similarity, based on high-level information that is only relevant for cross-linguistic comparisons, is at least partly responsible for disparities between intra- and inter-language effects of low-level similarity. In cases of conflict, phonemic similarity tends to override acoustic perceptual similarity, with the result that cross-linguistic speech patterns
often depart from predictions based on acoustic and/or perceptual similarities. As summarized in section 3, this trend is found in a range of cross-linguistic studies examining L2 perception, L2 production, L2-influenced phonetic drift in L1 production, and loanword adaptation.

Although I have argued that L2 users are swayed by phonemic correspondences when phonemic information is in conflict with low-level information, it is important to emphasize that this is a tendency, not a rule. In section 3.4, it was pointed out that loanword adaptation is not all about phonemic correspondences, and that the ultimate form of a loanword often bears traces of sensitivity to phonetic properties of the source language. For example, while Burmese adapts allophonically aspirated English voiceless stop allophones with unaspirated voiceless stops, Thai generally adapts these English allophones with aspirated voiceless stops rather than with unaspirated voiceless stops (Kenstowicz and Suchato 2006). The English-to-Thai mapping is thus an apparent counterexample to the ranking of phonemic similarity over acoustic phonetic similarity. However, when considered more carefully, the disparity between Burmese adaptation and Thai adaptation may actually be due to phonemic considerations after all. In Burmese, the adaptation of English voiceless stops with unaspirated voiceless stops allows aspirated voiceless stops (namely, /pʰ/) to be used to adapt certain English fricatives (namely, /f/) that are absent from the Burmese inventory, thus preventing phonemic contrasts between English fricatives and stops from being neutralized (Chang, 2012b). Thai also lacks certain English fricatives – namely, /θ/, which it adapts as /t/ (Kenstowicz and Suchato 2006). Adaptation of English voiceless stops with Thai unaspirated voiceless stops would, therefore, neutralize the contrast between English /θ/ and /t/, so instead they are adapted with Thai aspirated voiceless stops, which preserves the contrast between /θ/ and /t/.

In other words, although it is possible for “phonetic” kinds of similarity to prevail over “phonological” kinds of similarity in cases where they make different predictions, this appears to occur in extenuating circumstances having to do with other phonological considerations (or, alternatively, with an insufficient knowledge of the L2 phonology). As yet, it is not clear that an L2 user with phonemic knowledge of the L2 would ever weigh phonetic information at the expense of phonemic information (e.g., maximizing phonetic detail in a way that neutralizes phonemic contrast). The claim made here is that this is unlikely to happen because phonemic similarity has a privileged status stemming from its connection to establishing and maintaining meaningful linguistic contrast.

In section 3.1 it was observed that cross-linguistic similarity between segments differs from within-language similarity between segments in two ways: the relevance of between-system comparisons at a phonemic level, which are applicable only in cross-linguistic situations, and the language-specific nature of cross-linguistic perceptual similarity, which arises due to cross-linguistic differences in the landscape of L1 “perceptual magnets” (Kuhl and Iverson 1995) for unfamiliar L2 sounds. When L2 phonemic information is available, it exerts a powerful influence on cross-linguistic segmental mapping that can override conflicting information from acoustic phonetic similarity. In this sense, phonemic similarity constitutes one of multiple
factors that may mask effects of “raw” perceptual similarity between languages. As discussed by Shaw and Davidson (2011), recoverability and uniformity are other factors that may interact with perceptual similarity in determining the output of cross-linguistic production. The challenge for future cross-linguistic speech research will be to account for how much cross-linguistic differences in the grammatical effects of perceptual similarity have to do with variation in the “P-maps” of speakers of diverse languages (due to the perceptual warping caused by linguistic knowledge) vs. other impinging factors, such as abstract phonemic comparisons.

5 Conclusion

The research findings reviewed in this chapter suggest that the way in which L2 segments are related to L1 segments differs fundamentally from the way in which L1 segments are related to other L1 segments. I have argued that this disparity is rooted in a phonemic level of segmental comparisons that is only relevant between languages. Phonemic similarity distinguishes cross-linguistic phonological similarity from within-language phonological similarity because only judgments of cross-linguistic similarity can be influenced by between-system analyses of two phonologies. To the extent that such phonemic comparisons may depart from acoustic and allophonic comparisons, the availability of this high-level information can lead to the appearance that low-level information is being ignored, since high-level information is likely to prevail in cases of conflict.

Although the studies discussed in this chapter provide evidence for the privileged status of phonemics in determining overall phonological similarity between L1 and L2 segments, it is logical to expect differences between L2 learners, who have phonemic knowledge of the L2, and naïve non-natives, who do not. If we can assume, as implied by Flege (1995), that L2 sounds undergo automatic equivalence classification with L1 sounds, this suggests that at the onset of L2 learning L1 and L2 sounds must be linked on the basis of low-level information. It is clear that perceived cross-linguistic similarity based on this kind of low-level information is related to cross-linguistic behavior (e.g., Baker et al., 2002; Aoyama et al., 2004), but what remains unclear is how perceived similarity between L1 and L2 sounds changes over the course of L2 learning. This question motivates several interesting avenues of future research into the effects of L2 phonemic information over time and the manner in which a changing level of cross-linguistic linkage modulates L1-to-L2 and L2-to-L1 influence as an L1 talker acquires an L2 phonology.

Acknowledgments

I am thankful to the audience at the CUNY Phonology Forum (especially William Idsardi and Douglas Whalen) and to the editors of this volume, Charles Cairns and Eric Raimy, for their insightful feedback on this paper. Any errors are mine alone.
Notes

1 Of course, the term “hierarchical” can refer to a variety of systems, ranging from a strictly feed-forward system to one in which information flows freely between modules. While the precise nature of the relationship between phonetics and phonology lies outside the scope of this chapter, the crucial point is that phonology is privileged (higher-ranked) relative to phonetics.

2 I say “at least” because I have clearly omitted other dimensions of phonological similarity (e.g., articulatory/gestural similarity, aerodynamic similarity) in the interest of focusing on the contrast between low-level and high-level similarity.

3 I am conflating acoustic and auditory similarity here since the distinction is not important for the main contrast between “phonetic” and “phonological” kinds of similarity. However, it is important to note that acoustic and auditory similarity are in fact different. While auditory impression may broadly reflect the acoustics of the speech signal, auditory distances are not linearly related to acoustic distances because of basic aspects of hearing and auditory processing such as the pattern of frequency response of the basilar membrane in the inner ear. This fact does not affect the main argument; nevertheless, it should be borne in mind that although the phonetic distances referred to throughout this chapter are acoustic, the relevant distances in regard to perceptual similarity are really auditory.

4 The reason for supplementing phonological features with the notion of “relative phonetics” (by which I mean “relative position in a relevant phonetic dimension”; e.g., “long” end of the voice onset time dimension) is to ensure a common currency of comparison between L1 and L2 sounds, which feature specifications may not always provide. For example, stop laryngeal categories – whether “voicing” or “aspiration” categories – have been widely described phonetically in terms of the acoustic property of voice onset time, but phonologically in terms of at least two different features, [±voice] and [±spread glottis], depending on whether the contrast is one of voicing or aspiration. Relating stop types in a “voicing” language to those in an “aspirating” language via feature matching is, therefore, problematic; however, doing so in terms of relative phonetics is straightforward, since relative position in the voice onset time dimension is something that can be meaningfully compared for stop types in different kinds of languages.

5 Another prediction that follows from this hypothesis is that in case of a conflict between acoustic similarity and allophonic similarity, listeners will, depending on the nature of the task, be swayed by allophonic similarity over acoustic similarity in determining overall perceptual similarity between a pair of segments. That is to say, listeners whose native language contains a productive alternation only between a pair of phones that are relatively acoustically distinct (e.g., [s] and [θ]) are expected to perceive that pair of phones as more similar than a pair of phones that are acoustically closer (e.g., [f] and [θ]) but do not participate in such an alternation. This seems to be a reasonable prediction, but for reasons of space the discussion below is limited to conflicts between acoustic similarity and phonemic similarity.

6 Acoustic proximity in these studies has generally been measured in terms of distance in the first two or three vowel formants ($F_1–F_3$). However, there are limits to estimating acoustic proximity in these terms, since $F_1–F_3$, though sufficient as acoustic cues for distinguishing most vowels, are not the only determinants of vowel quality. Thus, it should be noted that inclusion of additional acoustic dimensions – especially the fundamental frequency ($f_0$) and the temporal trajectories of frequency components – would give a
Determining Cross-Linguistic Phonological Similarity Between Segments

fuller picture of acoustic proximity between vowels and may help account for perceptual assimilations to an L1 vowel that is not the closest to an L2 vowel as measured on the basis of $F_1-F_3$ alone.

Like all phonological contrasts, voicing contrast is associated with multiple acoustic cues besides VOT (e.g., $f_0$ in the adjacent vowel). However, there is evidence that English speakers rely on VOT as the primary cue to voicing (Abramson and Lisker 1985), suggesting that for English speakers perceptual similarity between L1 and L2 stop types is likely to closely follow their VOT characteristics.

Naturally, many of these cross-linguistic linkages between segments argued to be based on phonemic similarity could also be explained in terms of orthographic similarity (i.e., being spelled with the same graphemes). However, even when the L1 and L2 share the same alphabet, not all of the data can be explained in this way; see LaCharité and Paradis (2005, pp. 240–241) for extensive arguments against attributing cross-linguistic mappings to orthographic influence.

References


10
Contrast and Vowel Features
San Duanmu

1 Introduction

The goal of this study is to explore whether we can define the set of possible segments, or consonants and vowels, in the world’s languages. I shall follow a common assumption, called “feature bounding” by Clements (2009), that the set of possible segments is delimited by a set of features. For example, if there are N features, each having a binary value and able to combine freely with other features, there are $2^N$ possible segments. I shall not discuss all possible segments though. Instead, I shall focus on basic vowels. I begin with some preliminary questions: What are basic vowels? What are features and how are they determined? Do we have adequate data for the task? Can we compare sounds and features across languages?

1.1 Basic vowels

Basic vowels are those that involve lip rounding, the backness of the tongue, the height of the tongue, and the tongue root gesture. They roughly correspond to those in an IPA chart, without diacritic marks for nasalization, breathiness, creakiness, and so on.

Two questions require some discussion though. First, how is vowel length represented? There are three views. First, a long vowel is made of two short vowels. Second, a long vowel is distinguished from a short vowel by the feature [+long]. Third, long and short vowels differ in timing slots: a long vowel is linked to two timing slots and a short vowel is linked to one. I shall argue for the third position.
The second question concerns diphthongs and triphthongs. One approach treats them as single vowels. The other treats them as combinations of two or three vowels. In the first approach, Standard Chinese has 21 vowels (Lee and Zee 2003). In the second approach, Standard Chinese has five vowels (Duanmu 2007). The first approach makes little reference to syllable structure. For example, in Standard Chinese, [iau] rhymes with [au]. This means that [iau] can be decomposed into [i] plus [au]. In addition, a short vowel can be followed by a consonant but a diphthong cannot. This suggests that a diphthong is equivalent to two sounds. In general, we can achieve a simpler analysis of syllable structure and a smaller inventory of vowels, if diphthongs and triphthongs are treated as clusters.

It is an open question whether diphthongs (and triphthongs) can be treated as two (or three) vowels in all languages. For example, some English speakers of New York City pronounce the vowel in *bath* and *cab* as [æə] (Cairns, p.c. 2013), which has been called a “short diphthong.” However, it has been noted that the New York [æ] undergoes “tensing” in such an environment (Benua 1995). If [æə] is long, it can be treated as a regular diphthong.

1.2 Features

Features serve two purposes (Halle 1962). The first is to distinguish sounds that are contrastive. The second is to define natural classes of sounds.

For the first purpose, a feature represents a minimal contrast between two sounds. A contrast is a difference between two sounds that can distinguish words in a language. Consider the examples in (1).

(1) Contrast in English

<table>
<thead>
<tr>
<th>Minimal</th>
<th>[s] vs. [z]</th>
<th>sip vs. zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-minimal</td>
<td>[s] vs. [v]</td>
<td>set vs. vet</td>
</tr>
</tbody>
</table>

It can be shown that the difference between [s] and [z] is minimal, commonly known as [voice], in the sense that [voice] cannot be further divided into two (or more) components each of which can itself be contrastive. It can also be shown that the difference between [s] and [v] is not minimal, in the sense it can be divided into two components, [voice] and “place,” each of which can be contrastive by itself.

For the second purpose, there is an assumption that every natural class involves at least one feature value that is shared by all its members. For example, in English, the set of sounds that precedes the plural suffix [s] is a natural class and share the feature [-voice].

Ideally, features obtained from contrasts are the same as those obtained from natural classes. However, while linguists agree on what a contrast is, they do not always agree on what constitutes a phonological pattern or whether such a pattern always involves a natural class (see Mielke 2008). In this study, therefore, I focus on contrast only.
1.3 Adequacy of data

I shall use two databases: UPSID (Maddieson and Precoda 1990) and P-Base (Mielke 2004–2007). UPSID contains 451 phoneme inventories and P-Base contains 628. Compared with the total number of languages in the world today, estimated to be at 6,000 (Moseley 2010), the databases may seem small. It is, therefore, natural to ask whether they are adequate.

UPSID was compiled by selecting one language from each typological group. It is a reasonable representation of the world’s languages therefore. P-Base was compiled by collecting all inventories on which there is a published grammar book at two large university libraries (Ohio State University and Michigan State University). It is, therefore, also an objective representation of what we know.

Some linguists are optimistic that we already know enough. For example, Ladefoged and Maddieson (1996: 1–2) offer an upbeat statement below:

We believe that enough is now known to attempt a description of the sounds in all the languages of the world … The ‘global village’ effect means that few societies remain outside the scope of scholarly scrutiny. In all probability there will be a sharp decrease in the rate at which previously unknown sounds are drawn to the attention of phoneticians.

Besides the issue of coverage, other questions have been raised (Simpson 1999; Vaux 2009). For example, there are different treatments of diphthongs. Similarly, different analyses may choose different symbols to represent certain phonemes. For example, in P-Base, Spanish has [b d g], but in UPSID they are given as [β ð γ]. A further issue is typographic convenience, as noted by Ruhlen (1976). For example, the IPA symbol [a] is supposed to be a low front vowel and [a] a low back vowel, but when a language does not have both, [a] is often used for [a]. Such issues pose problems if we are interested in the frequencies or the markedness of sounds (Clements 2009), but not if we are interested in contrasts among different sounds. For example, whether the Spanish sounds are [b d g] or [β ð γ], we may need to distinguish all of them, if some languages have all of them. Similarly, whether a low vowel in a given language is [a] or [a], as long as some language contrasts [a] and [a], we can capture the distinction. Moreover, our method will identify all inventories that appear to contain unusual contrasts, each of which will then be examined manually.

1.4 Cross-language comparison

A more serious question is whether sounds and features can be compared across languages. For example, both German and Norwegian use the same vowel symbols [i y e ø], yet the German vowels are systematically higher (Disner 1983: 67). Such small but systematic differences between languages are quite common. What is the
reason to say that the Norwegian vowels are the same as those in German, beyond the fact that people happened to have chosen the same vowel symbols? Should such differences be distinguished at all?

Consider another common problem, illustrated with the backness of the tongue in (2), where A-D are four vowels in two hypothetical languages L1 and L2.

(2) Backness of four vowels A-D in two hypothetical languages L1 and L2

The point of interest is the analysis of C. In L1, A is front and B back. If we consider L2 alone, we may call C front and D back. Can we then say then that A is the same vowel as C? Phonetically, they are different. However, if no language has more than two degrees of backness, we can consider A and C to be the same, both being front. If, however, we find a language that has three degrees of backness, we may need to reconsider whether C is front or central.

Given such issues, at least three views have been offered. According to the first (e.g. Ladefoged 1972; Disner 1983; Port and Leary 2005), each language is different and should be analyzed on its own. It makes little sense to identify sounds in one language with those in another. For example, in L1 of (2), we can call A front and B back, and in L2 we can call C front and D back, but it makes little sense to identify A with C, because “front” in L1 does not mean the same as “front” in L2. This approach fails to address the maximal number of contrasts in each phonetic dimension. In addition, this approach fails to appreciate the possibility that a mapping relation can hold between sounds that are not phonetically identical, such as A and C in above, especially if no language has more than two degrees of contrast in backness, to be seen below.

According to the second view, sounds in different languages can be equated to each other, if we have a universal feature system (e.g. Chomsky and Halle 1968, for whom features are “substantive universals”). However, it remains to be shown how such a feature system is to be discerned from available inventory databases.

According to the third view, features can be derived from physical landmarks in the vocal tract (the “quantal theory” of Stevens 1972). Therefore, at least some features can be identified across languages. It is unclear though how many degrees of contrast such a theory predicts for each phonetic dimension and how well the predictions fare against available databases.

The goal of this study is to gather empirical evidence on the maximal number of contrast in each phonetic dimension. Once that is known, we shall have some idea of the maximal number of possible segments. Therefore, the present study should be of interest to all parties in the theoretical debate.
2 Method

Our method follows three guidelines. I call them the Principle of Contrast, Known Feature First, and Maxima First. They are given in (3)–(5).

(3) The Principle of Contrast
   a. If two sounds A and B can contrast in any language, they must be distinguished by at least one feature.
   b. If two sounds A and B never contrast in any language, they need not be distinguished by a feature.

(4) Known Feature First
   Unless evidence requires otherwise, use known features or properties first before introducing a new feature.

(5) Maxima First
   a. First, search through every language in order to determine the maximal number of contrasts in each phonetic dimension.
   b. Then, interpret the sounds in each language in terms of the maximal number of contrasts in each phonetic dimension.

2.1 Principle of Contrast

The Principle of Contrast is commonly used in the analysis of individual languages (e.g. International Phonetic Association 1999: 160). Our definition, however, extends it to cross-language comparisons. In particular, we can define allophones in terms of contrasts in other languages. This is shown in (6).

(6) Allophones: two sounds A and B (which have some phonetic similarity) are allophones of the same phoneme in a language if and only if
   a. A and B do not contrast in this language, and
   b. A and B contrast in another language.

Allophones are sounds that can potentially contrast in some language. If two sounds A and B never contrast in any language, they need not be distinguished as allophones in any language. For example, [m] (released) and [m˭] (unreleased) are sometimes listed as allophones in English. However, if they never contrast in any language, they need not be distinguished. Similarly, if [m ɱ] never contrast in any language, they need not be distinguished either.

There is clear evidence for the Principle of Contrast. For examples, consider (7), which shows eight vowels by two female speakers of American English, and (8), which shows three vowels (20 tokens each) by one female speaker of American English.
(7) \([i \, \varepsilon \, \ae \, \alpha \, \o \, \u]\) by two female speakers (solid line vs. broken line) of Midwestern American English, measured by the present author

![Diagram showing vowel space with labels](image)

(8) \([i \, \alpha \, u]\) (20 tokens each) by one female speaker of American English, based on the narrow transcription of speaker s0101a from Columbus, Ohio, in the Buckeye Corpus (Pitt et al. 2007; measured by San Duanmu)

![Diagram showing vowel space with labels](image)

In (7), we see that some corresponding vowel pairs are quite different phonetically, yet their differences are ignored, since neither speaker considered the other to have any accent. In (8), we see that even for the same speaker, what are heard as \([i \, u \, \alpha]\) by phonetic transcribers in fact vary a lot, which shows again that non-contrastive differences can be ignored.

Once we recognize the Principle of Contrast, we can use it for feature analysis. Consider Ao (Gowda 1991), which has \([i \, u]\), and Apatani (Abraham 1985), which has \([i \, i]\). The IPA symbols suggest that there are three degrees of backness: front \([i]\), central \([i]\), and back \([u]\), but neither language has a three-way contrast. If no other language has a three-way contrast, we can analyze the vowels in (9), where \([i]\) in Apatani is reinterpreted as \([u]\), and each language only has a two-way contrast in backness.
Analysis of backness in two languages

Ao (Gowda 1991)  \( \text{i} \rightarrow \text{u} \)
Apatani (Abraham 1985)  \( \text{i} \rightarrow \text{u} \)

2.2 Known Feature First

The purpose of Known Feature First is to minimize redundancy in representation. For example, consider the difference between \( \text{ə} \) and \( \text{ʌ} \) in English, shown in (10).

Representing the difference between \( \text{ə} \) and \( \text{ʌ} \) in English

| Feature difference | \( \text{ə} \) | central |
|-------------------|---------------|
| \( \text{ʌ} \)     | back          |
| Stress difference | \( \text{ə} \) | - stress |
|                   | \( \text{ʌ} \) | + stress |

When \( \text{ə} \) and \( \text{ʌ} \) are distinguished, \( \text{ə} \) appears to be central and \( \text{ʌ} \) back. However, \( \text{ə} \) is an unstressed vowel and \( \text{ʌ} \) a stressed one. Since the distinction is already represented by stress, there is no need to represent it again by a feature (of backness).

Similarly, consider “advanced tongue root” (ATR), “tense,” and “pharyngealized.” These features are similar in various ways. For example, among high vowels, tense correlates with advanced tongue root, or \(+\text{ATR}\), and lax correlates with retracted tongue root, or \(\text{-ATR}\), although the correlation is less obvious among low vowels (Halle and Stevens 1969). In addition, pharyngealized vowels (reported in !Xóó, Traill 1985) are made with retracted tongue root, or \(\text{-ATR}\). Therefore, unless ATR, tense, or pharyngealized vowels contrast with each other in some language, we may not need all three features.

As a third case, consider vowel height (or factors that affect vowel height). Two options are shown in (11).

Representing vowel height

<table>
<thead>
<tr>
<th>1 feature</th>
<th>2 features</th>
</tr>
</thead>
<tbody>
<tr>
<td>[i] high 1</td>
<td>+ high, +ATR</td>
</tr>
<tr>
<td>[i] high 2</td>
<td>+ high, -ATR</td>
</tr>
<tr>
<td>[ɛ] high 3</td>
<td>-high, +ATR</td>
</tr>
<tr>
<td>[e] high 4</td>
<td>-high, -ATR</td>
</tr>
</tbody>
</table>

If we use one feature, we need a four-way contrast. If we use two features, we need a two-way contrast each. There is evidence that some languages need two features, such as Kinande (Kenstowicz 2009). Unless there is evidence otherwise, we can use two features for other languages, too.
2.3 Maxima First

As discussed above, without knowing the maximal number of possible contrasts in a phonetic dimension, it is difficult to compare sounds across languages. Maxima First offers a solution by setting up a system of reference for cross-language comparisons. Maxima First interacts with the Principle of Contrast, in that only contrastive differences in each feature dimension are represented. Maxima First also interacts with Known Feature First, in that when a language seems to show a larger than expected number of contrasts in a feature, we need to examine whether the contrasts can be represented with two (or more) known features, so that the number of contrasts in the original feature is reduced.

2.4 Procedure

Given the discussion above, we adopt the procedure in (12), where step (12d) involves the use of the Principle of Contrast and Known Feature First, to be illustrated below.

(12) Procedure of vowel analysis
a. Extract a complete list of distinct vowel transcriptions.
b. Divide the list into a set of basic vowels (those involving backness, height, rounding, and ATR) and those that are made of a basic vowel plus one or more additional features.
c. Search through every language and extract inventories that seem to involve a controversial contrast (e.g., a three-way contrast in backness).
d. Reexamine each extracted inventory and see if alternative analyses are available.

3 Result

I illustrate the process with data from UPSID. UPSID contains 269 distinct vowel transcriptions, totaling 3,833 tokens. They are divided into eight categories in (13). The category “laryngeal” includes laryngeal (creaky) and breathy (murmur) vowels. The category “others” includes voiceless, retroflex, and fricative vowels.

(13) Vowels in UPSID

<table>
<thead>
<tr>
<th>Category</th>
<th>Type</th>
<th>Token</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>45</td>
<td>2,699</td>
</tr>
<tr>
<td>Diphthong</td>
<td>89</td>
<td>201</td>
</tr>
<tr>
<td>Long</td>
<td>40</td>
<td>287</td>
</tr>
<tr>
<td>Nasalized</td>
<td>30</td>
<td>508</td>
</tr>
</tbody>
</table>
Basic vowels refer to those that involve the backness of the tongue, the height of the tongue, lip rounding, and ATR (advanced tongue root, or tenseness) only. They are the most common vowels and cover 70% of all vowel tokens. Diphthongs and long vowels are composed of basic vowels and need not be discussed separately. Similarly, nasalized or laryngealized vowels need little elaboration, since features like nasalization, creakiness, and murmur can be added to a basic vowel. It can be shown, too, that the distinction between regular and “over-short” vowels is similar to that between long and short vowels, since no language has a three-way contrast among over-short, regular, and long vowels. Therefore, this category is not discussed either. The category “pharyngeal” is similar to ATR, where “pharyngealized” corresponds to [−ATR] and “non-pharyngealized” to [+ATR]; in addition, we found no language in which pharyngeal and ATR vowels contrast. Finally, the category “others” involves voiceless vowels, retroflex vowels, and fricative vowels. No voiceless vowel is found to contrast with a regular vowel. Retroflex vowels can be represented with a coronal feature added to regular vowel features. Fricative vowels are found not to contrast with syllabic fricatives. Therefore, in what follows, I discuss basic vowels only.

UPSID assumes seven degrees of height, three degrees of backness, and two degrees of rounding. This gives 42 possible basic vowels, of which 38 are found, shown in (14).

(14) Basic vowels in UPSID. Vowels not found are in parentheses.

```
    Front  Central  Back
  High     i  y  i  u  u  u
High (lower) i  y  i  ø  ø  ø
Mid (higher) e  ø  ø  ø  y  ø
Mid        e  ø  ø  ø  y  ø
Mid (lower) e  ø  ø  ø  ø  ø
Low (raised) æ (æʷ)  y (yʷ)  ø  ø
Low        a (aʷ)  a (aʷ)  ø  ø
```

Seven other basic vowels involve an additional diacritic and do not fit into (16). They are [e̞ e̞ ø̞ y̞ ø̞ ø̞ i̞], where [− + _] indicate retracted, fronted, and velarized, respectively. They are found not to contrast with [e ø y ø i] and so require no further discussion.

Many phonologists assume just two degrees of backness and two degrees of height for mid vowels. Therefore, the controversial aspects of (14) are three degrees of backness and three degrees of height for mid vowels. A search was made for all triplets in these two features. There are 12 triplets for backness, shown in (15).
(15) Twelve contrastive triplets in backness

\[ [i \ i \ \text{UI}], [i \ i \ \text{UI}], [\text{e} \ \text{e} \ \text{y}], [\text{e} \ \text{e} \ \text{y}], [\text{e} \ \text{e} \ \text{y}], [\text{a} \ \text{a} \ \text{a}] \\
\[ [y \ u \ u], [y \ o \ o], [\text{e} \ \text{e} \ \text{y}], [\text{a} \ \text{a} \ \text{a}]

A search through the inventories in UPSID yields three languages, with one triplet each. They are shown in (16).

(16) Search result for backness triplets

<table>
<thead>
<tr>
<th>Language</th>
<th>Triplet found</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moro</td>
<td>[e \ e \ y]</td>
</tr>
<tr>
<td>Nimboran</td>
<td>[i \ i \ UI]</td>
</tr>
<tr>
<td>Woisika</td>
<td>[a \ a \ a]</td>
</tr>
</tbody>
</table>

Having gathered the exceptions, we examine each inventory to see if the case is valid. The vowel inventory of Moro is shown in (17), as given in the original source.

(17) Vowel inventory of Moro (Mr. and Mrs. Black 1971)

\[ i \ u \ e \ \partial \ \gamma, \ o \ a \]

The triplet of interest is [e \ e \ y]. However, according to the source, [\partial] occurs in unstressed positions only, whereas other vowels occur in stressed positions. This means that [\partial] is not a full phoneme and that Moro does not have a three-way contrast in backness.

Next we consider Woisika, whose vowel inventory is shown in (18), as given in UPSID, and (19), as given in the original source (Stokhof 1979).

(18) Vowel inventory of Woisika as given in UPSID

\[ i, \ i: \ o, \ u; \epsilon, \ e: \ \partial, \ o; \ a: \ \partial, \ o; \]

(19) Vowel inventory of Woisika as given in the original source (Stokhof 1979)

\[ i, \ i: \ o, \ u: \epsilon, \ e: \ \partial, \ o: \ \ae: \ a: \ \partial: \]

The intended three-way contrast in backness is [a \ a \ a], but as shown in the original source, there is a length difference between the central and back vowels, omitted in UPSID. If the difference between [a:] and [a] is represented by length, there is no need to represent it again by backness (central vs. back). Therefore, Woisika has no three-way contrast in backness.
Finally, consider Nimboran, whose vowels are \([i \ iy \ e \ y \ a]\), all of which are unrounded (Anceaux 1965: 9). The backness triplet is supposed to be \([i \ iy]\). However, according to the source (Anceaux 1965: 13–15), for some speakers \([i]\) is “rather tense” and “backed,” whereas \([iy]\) is slightly lowered. This means that \([iy]\) could differ in tenseness or ATR, while both being back and high. Therefore, the Nimboran case does not seem compelling either.

Next, we consider three-way contrast in height among mid vowels. There are six such triplets, shown in (20). A search through UPSID yields two hits, shown in (21), both in the language Klao.

\[
\begin{align*}
\text{(20)} & \quad \text{Six contrastive triplets in height among mid vowels} \\
& \quad [\varepsilon \varepsilon \varepsilon], [\phi \phi \phi\phi], [\varepsilon \varepsilon \varepsilon], [\varepsilon \varepsilon \varepsilon], [\varepsilon \varepsilon \varepsilon], [\phi \phi \phi]
\end{align*}
\]

\[
\begin{align*}
\text{(21)} & \quad \text{Height triplets for mid vowels found in UPSID, both in Klao} \\
& \quad [\varepsilon \varepsilon \varepsilon], [\phi \phi \phi]
\end{align*}
\]

Klao has both oral vowels and nasal vowels. The oral vowels are shown in (22), as given in UPSID.

\[
\begin{align*}
\text{(22)} & \quad \text{Oral vowels in Klao, as given in UPSID} \\
& \quad \begin{array}{c}
\text{High} \\
\text{Mid (higher)} \\
\text{Mid} \\
\text{Mid (lower)} \\
\text{Low}
\end{array} \\
& \quad \begin{array}{c}
i \\
\varepsilon \\
e \\
\varepsilon \\
a
\end{array} \\
& \quad \begin{array}{c}
u \\
\phi \\
o \\
\phi \\
\phi
\end{array}
\end{align*}
\]

However, in the source (Singler 1979: 63), the vowels are described differently, as shown in (23), where \([e \ o]\) are [-ATR] (expanded pharynx) and \([\varepsilon \ \phi]\) are [-ATR].

\[
\begin{align*}
\text{(23)} & \quad \text{Oral vowels in Klao, as given in the original source (Singler 1979: 63)} \\
& \quad \begin{array}{c}
\text{High} \\
\text{Mid} \\
\text{Low}
\end{array} \\
& \quad \begin{array}{c}
i \\
e \varepsilon \\
\varepsilon
\end{array} \\
& \quad \begin{array}{c}
u \\
o \phi \\
a \phi
\end{array}
\end{align*}
\]

Of interest is the fact that \([\varepsilon \ \phi]\) are not mid but low vowels. Therefore, there is no three-way contrast in height among mid vowels.

In summary, we found no compelling case of three-way contrast in backness, or in the height of mid vowels. This is true of both UPSID and P-Base. Therefore, the maximal number of basic vowels is not 42 but 20, of which only 19 are found, shown in (24), where there are two degrees of rounding and backness, and at most five degrees of height.

\[
\begin{align*}
\text{(24)} & \quad \text{Basic vowels in Klao} \\
& \quad \begin{array}{c}
\text{High} \\
\text{Mid} \\
\text{Low}
\end{array} \\
& \quad \begin{array}{c}
i \\
e \varepsilon \\
\varepsilon
\end{array} \\
& \quad \begin{array}{c}
u \\
o \phi \\
a \phi
\end{array}
\end{align*}
\]
(24) Basic vowels in UPSID and P-Base

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>High (lower)</td>
<td>ɪ</td>
<td>ʊ</td>
</tr>
<tr>
<td>Mid</td>
<td>e</td>
<td>ø</td>
</tr>
<tr>
<td>Mid (lower)</td>
<td>ɛ</td>
<td>œ</td>
</tr>
<tr>
<td>Low</td>
<td>æ</td>
<td>ø</td>
</tr>
</tbody>
</table>

4 A close look at vowel height

According to Chomsky and Halle (1968), Kiparsky (1974), and others, there are five binary features for basic vowels, [back], [round], [high], [low], and [ATR]. The system yields 24 basic vowels, shown in (25).

(25) Basic vowels proposed by Chomsky and Halle (1968) and Kiparsky (1974)

<table>
<thead>
<tr>
<th></th>
<th>-round</th>
<th>+round</th>
<th>-round</th>
<th>+round</th>
<th>-round</th>
</tr>
</thead>
<tbody>
<tr>
<td>+high, -low</td>
<td>+ATR</td>
<td>i</td>
<td>ɣ</td>
<td>ɯ</td>
<td>u</td>
</tr>
<tr>
<td>-ATR</td>
<td>ɪ</td>
<td>ɣ</td>
<td>ʊ</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>-high, -low</td>
<td>+ATR</td>
<td>ɛ</td>
<td>ø</td>
<td>ɤ</td>
<td>o</td>
</tr>
<tr>
<td>-ATR</td>
<td>ɛ1</td>
<td>ø</td>
<td>ɔ</td>
<td>ɔ1</td>
<td></td>
</tr>
<tr>
<td>-high, +low</td>
<td>+ATR</td>
<td>æ</td>
<td>æ2</td>
<td>ʌ</td>
<td>ɔ2</td>
</tr>
<tr>
<td>-ATR</td>
<td>æ</td>
<td>æ</td>
<td>ø</td>
<td>ø</td>
<td></td>
</tr>
</tbody>
</table>

The system is more parsimonious than most others, such as that of a standard IPA table (International Phonetic Association 1999) or that of UPSID. Still, the system exceeds what is needed to represent all vowel contrasts in UPSID and P-Base. If we use the same features, we can represent our results in (26), where some vowel symbols are slightly adapted and [ATR] is not used for low vowels.

(26) Basic vowels proposed by Chomsky and Halle (1968) and Kiparsky (1974), revised

<table>
<thead>
<tr>
<th></th>
<th>-round</th>
<th>+round</th>
<th>-round</th>
<th>+round</th>
<th>-round</th>
</tr>
</thead>
<tbody>
<tr>
<td>+high, -low</td>
<td>+ATR</td>
<td>i</td>
<td>ɣ</td>
<td>ɯ</td>
<td>u</td>
</tr>
<tr>
<td>-ATR</td>
<td>ɪ</td>
<td>ɣ</td>
<td>ʊ</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>-high, -low</td>
<td>+ATR</td>
<td>ɛ</td>
<td>ø</td>
<td>ɤ</td>
<td>o</td>
</tr>
<tr>
<td>-ATR</td>
<td>ɛ</td>
<td>ø</td>
<td>ɔ</td>
<td>ɔ</td>
<td></td>
</tr>
<tr>
<td>-high, +low</td>
<td>æ</td>
<td>æ</td>
<td>ø</td>
<td>ø</td>
<td></td>
</tr>
</tbody>
</table>

While (26) is simpler, the need for three degrees of height has not been demonstrated in our study. In addition, there are several other problems. First, there is a missing vowel in (26). Second, height seems to be the only phonetic dimension that
has three degrees of contrast (even though it is represented with two features). Third, it remains unclear why [ATR] does not apply to low vowels. Fourth, while minimally contrastive pairs between high and mid vowels are easy to find, minimally contrastive pairs between mid and low vowels are quite rare (see English examples below). Finally, there are languages where vowels fall into just two groups of height, even when they seem to show three degrees of height phonetically. This is the case in Turkish, whose vowels are shown in (27), where [a] is low and central.

(27) Vowels in Turkish (Zimmer and Orgun 1992: 44)

\[\begin{array}{c}
\text{i} & \ast & \ast \\
\text{ɯ} & \ast & \ast \\
\text{u} & \ast & \ast \\
\text{y} & \ast & \ast \\
\text{e} & \ast & \ast \\
\text{a} & \ast & \ast \\
\end{array}\]

However, with regard to vowel harmony, Turkish vowels fall into two degrees of height, shown in (28), where [a] is back and belongs to the same height category as phonetically mid vowels, both called “open” (Lewis 1967).

(28) Feature analysis of Turkish vowels (Lewis 1967: 14)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrounded</td>
<td>Rounded</td>
<td>Unrounded</td>
</tr>
<tr>
<td>Close</td>
<td>i</td>
<td>y</td>
</tr>
<tr>
<td>Open</td>
<td>e</td>
<td>o</td>
</tr>
</tbody>
</table>

It is worth asking, therefore, whether fewer basic vowels are sufficient to account for the inventories in UPSID and P-Base. If we only assume two degrees of height, the number of basic vowels is 16. This is shown in (29).

(29) Inventory of basic vowels in a two-height analysis

<table>
<thead>
<tr>
<th>[-back]</th>
<th>[+back]</th>
<th>[-round]</th>
<th>[+round]</th>
<th>[-round]</th>
<th>[+round]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+high]</td>
<td>[+ATR]</td>
<td>i</td>
<td>y</td>
<td>ɯ</td>
<td>u</td>
</tr>
<tr>
<td>[-ATR]</td>
<td>i</td>
<td>y</td>
<td>ɯ</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>[-high]</td>
<td>[+ATR]</td>
<td>e</td>
<td>o</td>
<td>y</td>
<td>o</td>
</tr>
<tr>
<td>[-ATR]</td>
<td>æ</td>
<td>œ</td>
<td>a</td>
<td>o</td>
<td></td>
</tr>
</tbody>
</table>

The choice of the phonetic symbols is flexible, but it has little consequence for our discussion. Let us now consider whether a 16-vowel system is sufficient to account for all vowel inventories. We shall look at British English, German, Swedish, and !Xóó. The first three are chosen because their patterns are well known and their
vowel inventories are fairly large. !Xóô is chosen because it has the largest vowel inventory in P-Base. We shall focus on whether (29) offers enough positions, rather than which position each vowel should go into.

British English has ten monophthongs. A possible analysis is shown in (30), following the transcription of Ladefoged (2001: 29). It is worth noting that there is phonological evidence for vowel length in English, both in syllable structure (Borowsky 1986) and in stress assignment (Halle and Vergnaud 1987; Hayes 1995).

(30) Two-height analysis of monophthongs in British English

\[
\begin{array}{c}
i: & \text{u:} \\
\text{i} & \text{u} \\
\varepsilon & \text{ɛ} \\
\text{æ} & \text{a} \\
\end{array}
\]

The two-height system has more than enough slots to accommodate English vowels. The vowels [a: a] can share the same slot, because they differ in length. The analysis shows that [tense] does not always correspond to [ATR]. Instead, length is a better correlate of [tense].

Next we consider German, which has 19 vowels (Wiese 1996; Kohler 1999, Fox 2005). They can be analyzed in (31), excluding [ə], which occurs in unstressed syllables only, and three diphthongs.

(31) Two-height analysis of stressed monophthongs in German

\[
\begin{array}{c}
i: & \text{y:} & \text{u:} \\
\text{i} & \text{y} & \text{u} \\
\varepsilon, \varepsilon: & \text{ɛ} & \text{a} \\
\end{array}
\]

[ɛ ɛ:] and [a a:] show again that [ATR] is independent from length. In addition, as in English, there is phonological evidence for vowel length in German (Giegerich 1985).

Next we consider Swedish, which has 17 vowels. They can be analyzed in (32), according to the transcription of Engstrand (1999: 141).

(32) Two-height analysis of Swedish vowels

\[
\begin{array}{c}
i: & \text{y:} & \text{u}, \text{u} \\
\text{i} & \text{y} & \text{u} \\
\varepsilon, \varepsilon: & \text{ɛ} & \text{a}, \text{a} \\
\end{array}
\]

Some positions are again filled by two vowels each, which differ in length, and there is no need to distinguish them by another feature.

Finally, let us consider !Xóô, which has 44 vowels, the largest vowel inventory in P-Base. The vowels are shown in (33).
Nasalization, murmur, and glottalization are non-basic features. In addition, there is no contrast between pharyngealization and ATR, in that pharyngeal vowels can be seen as [-ATR] and other vowels as [+ATR]. Thus, !Xôô has just eight basic vowels, analyzed in (34).

(34) Two-height analysis of basic vowels in !Xôô

\[
\begin{array}{l}
i \\
u \\
e \\
a \end{array}
\]

How many languages might have more than 16 basic vowels? To find out, consider the sizes of all vowel inventories in P-Base, shown in (35).

(35) Vowel inventory sizes in P-Base

<table>
<thead>
<tr>
<th>Vowel inventory size</th>
<th>Raw</th>
<th>Basic vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 16</td>
<td>595</td>
<td>627</td>
</tr>
<tr>
<td>Larger than 16</td>
<td>33</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>628</td>
<td>628</td>
</tr>
</tbody>
</table>

Out of 628 inventories, just 33 appear to have more than 16 vowels. If we exclude non-basic features (such as length, nasalization, murmur, etc.), just one language, Turkana, has more than 16 vowels. The Turkana inventory is shown in (36).

(36) Turkana vowel inventory: \{i u \i u \i u e o e e e e a a\}

In Turkana, there are just nine basic vowels, the rest being their voiceless counter-parts. The source, Dimmendaal (1983), notes that voiceless vowels lack stress. In a later analysis, Dimmendaal (1993: 131) no longer includes them in the phonemes inventory.

In summary, there is no compelling evidence for three degrees of vowel height, and 16 basic vowels seem sufficient to account for all inventories in P-Base and UPSID.

5 Concluding remarks

The goal of this study is to determine the range of possible segments in the world's languages, using UPSID and P-Base as our data. I have shown that, as far as basic vowels are concerned, only four binary features are needed: [back], [high], [round],
and [ATR]. This system yields a total of 16 basic vowels, shown in (29). In other words, there are far fewer possible vowels that commonly assumed, and far fewer features to distinguish them.

Our discussion has focused on contrast, with only occasional references to natural classes. With fewer features than before, our system shall be unable to account for certain natural classes. How then should the two systems, one based on contrast and one on natural classes, be reconciled?

It is worth noting that a contrast-based system is a minimal system that all phonologists would recognize. The question is whether we should expand the system in order to account for natural classes. The answer is not obvious, but I shall offer some speculations.

One possibility is that the contrast-based system should be expanded with additional features to account for natural classes. Such additional features could serve as potential contrasts, too, which happen not to be used in the languages we examined.

The other possibility is that natural classes not definable with contrast-based features need to be reexamined. This is a strong claim, but not entirely new. For example, in a large-scale study on natural class-based features, Mielke (2008) reports that 30% of what are thought to be natural classes are “unnatural,” in the sense that they cannot be defined by well-known features. Clearly, much work is needed, and it will be interesting to find out how to resolve the differences between evidence from contrast and that from natural classes.

Acknowledgments

I would like to thank Eric Raimy and Charles Cairns for hosting the CUNY Phonology Conference on the Segment, where this chapter was presented. I also thank the audience for their comments. I am also grateful to additional comments by Eric Raimy and Charles Cairns on a draft copy of this chapter, which led to much improvement in the revision.

References


11

The Phonetics and Phonology of Segment Classification

A Case Study of /v/

Christina Bjorndahl

1 Introduction

Regardless of one’s stance toward the existence and theoretical usefulness of the segment, cross-linguistic descriptions of inventories, phonological processes and phonotactics are inextricably bound up with some (presumably atheoretic) notion of segments. Statements such as “[s] is the most common fricative” rely on linguists having a common understanding of some kind of [s]-like entity that, regardless of the myriad ways that the segment we denote as /s/ can differ across languages, is reasonably construed as “the same.” Nevertheless, a precise formulation of what it is that allows such statements has been elusive: the “same” segment in two different languages can differ in terms of articulatory configuration, acoustic realization, distribution, and/or the phonological processes in which it participates. That the “s” in one language is not, in fact, the same as the “s” in another, at any level of analysis, is not simply a perversely philosophical observation, but rather cuts to the very core of linguistic theorizing, data collection and analysis. Any cross-linguistic comparison of segments is plagued by the possibility that the entities under consideration are not, in fact, the same, and hence not comparable. Nevertheless, even without a rigorous notion of the segment, a greater need for explicitness in the domain of comparison is required if we are to understand how segments differ, and what similarities underlie these differences.

The problem of cross-linguistic segment identity is closely tied to issues of segment classification. A standard approach is that sounds are classified based on their phonetic attributes. For example, the sound [z] is produced with vocal fold vibration, hence is voiced, and so is classified by attributing to it the feature [+voice]; such a featural specification classifies [z] together with other sounds such as [b, d, f, y]. Voiced sounds stand in opposition to [-voice] sounds, such as [p, t, s, x], and pairs
The Phonetics and Phonology of Segment Classification

237

such as [z, s] or [b, p] are typically considered a pair with respect to the feature [voice]. That this is so can be seen in how most consonant inventories are presented, with the voiced segment on the right and the voiceless one on the left, as is, for example, the representation of consonants in the IPA. Attributing a phonological feature (in this case, [voice]) on the basis of the phonetic properties of the phone (in this case, vocal fold vibration) relies on there being a fairly straightforward mapping between the phonological identity of a segment and its phonetic realization. This chapter probes this assumption by asking whether a segment’s phonological identity and classification can be inferred from the phonetic properties of its phone, focusing on the segment transcribed as /v/.

A frequent, if tacit, assumption is that when a consonant inventory contains both /v/ and /f/, they form a voiced‐voiceless obstruent pair, much like /z/ and /s/. This is indeed the case for Greek, in which /v/ distributes as the voiced counterpart to voiceless /f/: for example, both /v/ and /f/ are subject to the general requirement that obstruents agree in voicing, hence [evyłotos] “eloquent” vs. [efstaθia] “steadiness” (same prefix). However, the general assumption is challenged by languages in which /v/ does not pattern as the voiced counterpart to /f/. In Serbian, for example, /v/ has the distribution of a sonorant: /v/ can follow both voiced and voiceless obstruents, yielding contrasts such as [tvoj] “your” vs. [dva] “two,” flouting the otherwise robust phonotactic constraint that obstruent clusters must agree in voicing, a constraint that /f/ obeys without exception. The most striking case is Russian, in which /v/ shares the distribution of a sonorant, as in Serbian, but patterns ambiguously with respect to voicing processes. Like obstruents, Russian /v/ undergoes final devoicing and regressive voicing assimilation, devoicing in both cases to [f], but like sonorants fails to trigger regressive voicing assimilation. In sum, /v/ patterns as a sonorant in pre‐sonorant position, and thus voicing contrast in obstruents is maintained before /v/, yielding [tver] “Tver” vs. [dver] “door”; in pre‐obstruent position, /v/ patterns as an obstruent, hence /v supe/ “in the soup” is realized as [fsupe]. Word‐finally, /v/ also patterns as an obstruent, and the voicing contrast between /v/ and /f/ is neutralized word‐finally, hence [prava] and [praf] “right (fem./masc.).”

Striking as the Russian case is, it becomes even more interesting when viewed in light of the typology presented above. In some languages, exemplified by Greek, /v/ patterns as the voiced obstruent corresponding to /f/ and hence is classified as [+sonorant]; in other languages, such as Serbian, /v/ and /f/ do not comprise a voicing pair and /v/ receives the phonological classification of [‐sonorant]. Finally, in Russian, the patterning of /v/ is ambiguous with respect to whether it is classified as either [+sonorant] or [‐sonorant]. The distribution and patterning of Russian /v/, paralleled in languages as diverse as Hungarian and Hebrew, is a puzzle for phonological theory (Barkai and Horvath 1978; Hayes 1984; Jakobson 1978; Kiparsky 1985; Kiss and Bárkányi 2006): if /v/ is specified as an obstruent, then it should trigger regressive voicing assimilation, and if it is specified as a sonorant, it should not undergo voicing assimilation or final devoicing. In the words of Jakobson (1978), “the Standard Russian V… occupies an obviously intermediate position between the obstruents and the sonorants.”
The classificatory typology of /v/ with respect to the feature [sonorant] is based on phonological patterning, and one might ask how the phonetic attributes of [v] align with the phonological classification. Phonetically, the classificatory cues of [v] with respect to the obstruent-sonorant divide are not as robust as those of other obstruents. The aerodynamic constraints on voicing and frication conflict: insofar as voicing is maintained, the degree of frication is reduced. Ohala (1983) comments on the entire class of sounds that [v] belongs to, the nonsibilant voiced fricatives:

For the sake of continued voicing the oral pressure should be low, but for the sake of frication the oral pressure should be high, that is, the difference between oral pressure and atmospheric pressure should be high enough to cause high air velocity through the consonantal constriction. Meeting both of these requirements simultaneously may be difficult. To the extent that the segment retains voicing it may be less of a fricative, and if it is a good fricative it runs the risk of being devoiced … The noise component … on nonsibilant voiced fricatives (\[β, v, θ, j, y, ð\]) is often so weak as to be barely detectable. (p. 202)

We may further ask whether there exists a correspondence between the phonological classification of /v/ and its phonetic realization. If the phonetic attributes of [v] tokens in Greek, Russian and Serbian are all significantly different from each other, then this supports the idea that there may be a fairly strong correlation between phonological status and phonetic realization, and this scenario is the simplest from a classificatory standpoint. Another possibility is that the correlation is stronger for some languages than for others. Finally, it might be that the phonetic attributes of [v] in the three languages are not distinguishable, suggesting that the relationship between the phonological status and phonetic realization might be highly abstract.

In section 3 I present the results of an acoustic study designed to establish the phonetic correlates of [v] in Greek, Russian, and Serbian, which exemplify three different cases of phonological patterning of /v/. First though, I review the phonological arguments for /v/’s classification in section 2, as well as various issues crucial to establishing a phonological typology of segment classification.

## 2 Phonological classification of /v/

This chapter presents evidence for the phonological classification of /v/ in Greek, Serbian, and Russian. As a voiced, labiodental fricative, /v/ is expected to distribute and pattern as an obstruent, but the typology of its phonological patterning appears to be decidedly richer.

Table 11.1 summarizes the typology of /v/’s phonological status with respect to the obstruent-sonorant divide, distinguishing between its distribution and patterning. In Greek, /v/ distributes as an obstruent, reflecting its historical status as the voiced bilabial plosive /b/; in the Slavic languages /v/ was historically a sonorant /w/, reflected in its distribution in both Russian and Serbian. In Greek and Serbian, the
The Phonetics and Phonology of Segment Classification

239

parcelling of /v/ parallels its distribution, in contrast to Russian where /v/ patterns both with sonorants and with obstruents. I have yet to find a language in which /v/ distributes as an obstruent but has ambiguous patterning.

In order for the comparison between Greek, Russian, and Serbian to be robust, inventory structure was controlled for as much as possible. Thus each language has both /v/ and /f/, but none of the three languages has a labial approximant such as /w, ʋ, β̞/ against which /v/ might contrast (either phonologically or phonetically).

2.1 Greek

Phonological evidence shows that Greek /v/ is an obstruent. The Greek consonant inventory contrasts voiced and voiceless fricatives at four places of articulation, as seen in Table 11.2. Remarkably, three of these four pairs are non-strident, a state of affairs with a historical explanation: the voiceless fricatives /f, θ, x/ derive from Ancient Greek aspirated stops /pʰ, tʰ, kʰ/; the voiced fricatives /v, δ, γ/ derive from voiced stops /b, d, g/. The modern voiced stops /b, d, g/ arise from an underlying nasal followed by the corresponding homorganic stop, a pronunciation that is maintained post-vocally, and in fact how these sounds are rendered orthographically. The voiced non-strident fricatives /v, δ, γ/, together with their voiceless counterparts, thus derive historically from obstruents.

The distribution of [v] parallels that of the voiced non-strident fricatives [δ, γ], and this entire class mirrors the patterning of the voiceless non-stridents [f, θ, x]. Voiced and voiceless non-stridents combine with other non-stridents and sibilants with the same voicing specification, thus [fθ, fx, xθ, vδ, vγ, γδ] and [sf, sθ, sx, zv, zγ]; obstruent clusters must agree in voicing, thus [*sy, *zf]. Obstruents of either voicing specification may precede liquids, thus [pl, pr, bl, br], a pattern that extends to both the voiced and voiceless non-stridents, thus [fl, fr, vl, vr, θr, δr, xl, xr, γl, γr]. Like [f, θ, x], the voiced non-stridents cannot follow either a liquid or nasal in an onset cluster, thus [*lf, *lv, *lθ, *lδ]. Three-consonant clusters are irrelevant for distinguishing the distribution of [v] from other obstruents, as all such clusters are composed of [s], followed by a voiceless obstruent, followed by a liquid or nasal; none of the voiced non-stridents [v, δ, γ] can appear in such clusters in any position.

Within a word, across morphological boundaries, obstruent clusters take on the voicing of the rightmost obstruent; though these morphemes are no longer

<table>
<thead>
<tr>
<th>Table 11.1</th>
<th>Typology of the phonological status of /v/.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greek</td>
<td></td>
</tr>
<tr>
<td>Distribution</td>
<td>Obstruent</td>
</tr>
<tr>
<td>Patterning</td>
<td>Obstruent</td>
</tr>
<tr>
<td>Status</td>
<td>Obstruent</td>
</tr>
</tbody>
</table>
productive, (1) and (2) show that [v] participates fully in this process, as both a trigger and a target.

(1) [evɣlotos] ‘eloquent’ vs. [efstaθia] ‘steadiness’ (prefix /ev-/)
(2) [dizvatos] ‘rough’ vs. [disforia] ‘discomfort’ (prefix /dis-/)

The evidence for regressive voicing assimilation applying across word boundaries is limited because Greek words may only end in /s/ or /n/ (with few exceptions). Therefore it is only possible to test whether /v/ is a trigger for regressive voicing assimilation; as seen in (3)–(5), all voiced segments (both voiced obstruents and sonorants) trigger voicing of /s/ in these cases.

(3) /tusɣambrus/ [tuzɣambrus] ‘the grooms’ (acc.)
(4) /tis mamas/ [tiz mamas] ‘the mother’s’
(5) /as valis/ [az valis] ‘you may put’

### Table 11.2 Consonant inventory of Greek.

<table>
<thead>
<tr>
<th>Labial</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>p b t d</td>
<td>k g</td>
<td></td>
</tr>
<tr>
<td>Affricate</td>
<td>ts dz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>f v θ δ s z x γ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td></td>
<td></td>
<td>l</td>
</tr>
<tr>
<td>Rhotic</td>
<td></td>
<td></td>
<td>r</td>
</tr>
</tbody>
</table>

2.2 Serbian

In Serbian, /v/ clearly patterns as a sonorant. The consonant inventory of Serbian is presented in Table 11.3, in which I classify /v/ as an approximant to reflect the phonological facts discussed in this section.

Like Greek, Serbian has /f/, but the historical roots of /v/ and /f/ in Serbian differ markedly. Old Church Slavonic had /v/, which is often rendered as /w/; /f/ entered through borrowings, mostly from words of Greek origin, patterning phonologically as an obstruent. This situation is mirrored in Serbian, as /f/ is found mostly in borrowings, and /v/ continues to have the distribution of a sonorant.

Like sonorants, /v/ need not agree in voicing with a preceding obstruent, thus [pl, bl, tv, dv]; however, obstruent clusters must agree in voicing, thus [pt, bd, st, zd], but [*pd, *zt]. As the first member of a consonant cluster, /v/ can only precede the liquids /l, r/; in this, it patterns like the other labial sonorant /m/. Reflecting its status as a loan segment, /f/ does not readily enter into consonant clusters, appearing only in [sf] as the second member, and as the first member in [fl, fr, fj].
Serbian does not have final devoicing. The voicing specification of obstruent clusters is determined by the rightmost obstruent, as seen in (6) to (8); /v/ does not trigger voicing assimilation, as in (9). In (10) and (11) we see that /v/ also does not devoice under voicing assimilation.

\begin{align*}
(6) & /s-paziti/ [spaziti] 'observe' \\
(7) & /s-gaziti/ [zgaziti] 'trample' \\
(8) & /s-loʒiti/ [sloʒiti] 'put together' \\
(9) & /s-variti/ [svariti] 'digest' \\
(10) & /sav-a/ [sava] 'Sava' \\
(11) & /sav-ka/ [savka] 'Sava (dim.)'
\end{align*}

2.3 Russian

The Russian consonant inventory is presented in Table 11.4. Like Serbian, /v/ in Russian derives from a historical sonorant, reflected in its distribution, and /f/ entered the lexicon through borrowings. However, unlike Serbian, as /v/ devoices to [f] word-finally and under regressive voicing assimilation, [f] is better integrated into Russian phonology than it is in Serbian.

Distributionally, Russian /v/ patterns with sonorants. First, like the sonorants /m, n, l, r, j/, /v/ may follow either voiced or voiceless obstruents in two-consonant onset clusters, thus [pl, bl, tv, dv]; moreover, such clusters are particularly common, as Padgett (2002) noted. Of the three-consonant clusters permitted, they are typically comprised of an alveolar sibilant [s, z], a stop obstruent and always one of [n, l, r, j, v], as in [skvaʒina] “chink”. This and all other Russian examples are taken from Padgett (2002).

In terms of phonological processes, Russian /v/ patterns with both sonorants and obstruents. Obstruents in Russian participate fully in the voicing processes of final devoicing and regressive voicing assimilation. As shown in (12)–(14), underlyingly voiced word-final obstruents are devoiced, while sonorants are not; like obstruents, /v/
undergoes final devoicing and is realized phonetically as [f] in word-final position. Examples (15) through (17) show that the rightmost obstruent in a cluster determines the voicing of the entire cluster. Like obstruents, /v/ devoices before voiceless obstruents, but like sonorants, fails to trigger voicing assimilation, yielding contrasts such as [dv] vs. [tv].

| (12) | [sled-a] | [slet] | 'track (gen. / nom. sg.)’ |
| (13) | [mil-a] | [mil], *[mil] | 'dear’ |
| (14) | [prav-a] | [praf] | 'right (fem. / masc.)’ |
| (15) | /ot-pustit/ | [otpustit’] | 'release’ |
| (16) | /ot-brosit/ | [odbrosit’] | 'throw aside’ |
| (17) | /ot-nesti/ | [otnesiti] | 'carry away’ |
| (18) | /ot-vesti/ | [otvesti] | 'lead away’ |
| (19) | /v ruke/ | [v ruke] | 'in one’s hand’ |
| (20) | /v gorode/ | [v gorode] | 'in the city’ |
| (21) | /v supe/ | [f supe] | ‘in the soup’ |

These data show that despite its distribution as a sonorant, Russian /v/ patterns with both sonorants and obstruents in the active phonological processes.

3 Acoustic study of /v/

The purpose of the phonetic study reported here is to determine whether there exists a correlation between the phonological status of /v/ and its phonetic realization in Greek, Serbian, and Russian. Specifically, this study tests whether acoustic measures indicative of frication differ significantly between the three languages. If there is a one-to-one correspondence between phonological status and phonetic
realization then, all else being equal, tokens of Greek /v/ should be produced with the greatest amount of frication, tokens of Serbian /v/ with the least amount of frication, and tokens of Russian /v/ should be produced with a degree of frication somewhat intermediate between the two. Of course, “all else being equal” can be particularly elusive in a cross-linguistic phonetic study, and so it is worth discussing the measures taken to ensure that the comparisons are valid. First, as already explained, Greek, Serbian, and Russian were selected because they share the same local inventory with respect to /v/, all of them having /f/ but no labial approximant such as /w/ against which /v/ can contrast. Second, all three languages have a five vowel system (/i, e, a, o, u/), and only words with either /a/ or /o/ adjacent to the target segment were selected, minimizing coarticulatory effects such as palatalization triggered by high or front vowels. Finally, as discussed in detail in 3.2, the data are analyzed relationally: rather than compare the measures of /v/ directly across languages, the similarity of /v/ to /f/ is compared. The intuition underlying this approach is that in a language where /v/ patterns as an obstruent, as in Greek, it is the voiced counterpart to /f/, while in a language like Serbian, /v/ and /f/ do not comprise a voicing pair. By comparing the relationship between segments, much of the uncertainty present in cross-linguistic phonetic comparisons is removed.

3.1 Method

Seven native speakers of Greek (4F, 3M; aged 26–32), Serbian (3F, 4M; aged 29–47), and Russian (4F, 3M; aged 22–73) were recorded. None reported any hearing loss or showed any evidence of a speech impairment. All had left their home country after the age of 16 and the majority had been in North America for ten years or fewer at the time of recording. Dialect was partially controlled for, in that certain dialects were avoided, but it was not possible to find enough speakers from only one dialect. For all languages, a linguist native speaker was consulted regarding the dialect situation.

The recording session consisted of reading a word list in a frame sentence, with five randomized repetitions. Recordings took place in a sound attenuated chamber in the Phonetics Laboratory at either Cornell University or the University of Toronto. The recording device in both locations was a Sony SD722 digital recorder, and recordings were sampled at 44.1kHz with 16-bit quantization. At Cornell, the microphone used was an Electrovoice RE20 dynamic cardioid microphone; at Toronto, a DPA 4011 cardioid shotgun microphone for all speakers except SeM1, for which a Shure SM10-A head worn microphone was used.

All subjects except for GrF1, SeF1, and SeM3 were naïve as to the purpose of the experiment (GrF1, SeF1, and SeM3 are linguists; GrF1 and SeM3 helped design the word lists for Greek and Serbian, respectively). A few of the participants had some linguistics background, but when they were queried as to what they thought the
recording session was about, none guessed the purpose of the study correctly. Subjects were asked to read at a comfortable, conversational pace, to skip words they did not know, and were asked to repeat themselves if some disturbance occurred during a particular token (e.g., a cough, rustling of papers, etc.). Subjects were given breaks between reading each list, and each recording session took between 40 and 60 minutes. All subjects were remunerated $10 for their participation. The recorded signals were hand segmented in PRAAT, based on visual inspection of the waveform and spectrogram. The signal was then resampled to 22050Hz for processing and analysis in Matlab.

The segments /f, v, s, z, m/ were elicited in word-initial stressed syllables of disyllabic words. The additional segments were recorded to provide a comparison with /v/ at both ends of the obstruent–sonorant spectrum. Words were read in a frame sentence, given in (22) through (24), and in all cases, the vowel preceding the target word ended in /a/. The words used are listed in Table 11.5.

(22) Greek
eɣrapsa __________ tris fores
I wrote __________ three times

(23) Russian
sveta __________ skazala __________ odin ras
Sveta said __________ one time

(24) Serbian
kaʃe __________ jetsa __________ opet
said, Jetsa __________ again

3.2 Measures

In order to quantify the degree of frication, two spectral moments – the spectral centroid and skewness – were measured on the high-pass filtered signal. Such measures have been extensively used for fricative discrimination (see Jongman et al. 2000 for review), but under different conditions, and for different purposes.
The centroid is a measure of the location of energy concentration in the frequency domain, while skewness measures the overall asymmetry of the distribution. Voicing and noise will have opposite effects on the value of the centroid. Since noise is energy in the high-frequency range, higher centroid values correspond to noisier sounds. In contrast, voicing introduces energy in the low-frequency range, and thus voiced sounds will have lower centroid values than their voiceless counterparts. This is similarly true for skewness. Qualitatively, a negative skew indicates that the bulk of values lie to the right of the mean, while a positive skew indicates that the bulk of values lie to the left of the mean. Noisiness skews the energy distribution to higher frequencies (i.e., to the right), while voicing skews the energy distribution to lower frequencies (i.e., to the left). Nevertheless, as statistical measures, the spectral moments only make sense for a roughly unimodal distribution. For voiceless sounds, calculating the spectral moments over the whole distribution is unproblematic, as the distributions are unimodal to begin with, but for voiced fricatives with one peak in the lower frequency range and another peak in the high-frequency range, this criterion is not met. In order to circumvent this issue, and because we are interested in quantifying the degree of frication, the acoustic signal for all segments was first high-pass filtered to remove the effect of voicing and the first several harmonics, allowing an analysis of the distribution of energy in the high-frequency range.

The signal first underwent a 1500Hz high-pass fourth-order Butterworth filter. The spectral moments were then calculated on a moving 20-millisecond Hann window with 10-milliseconds overlap over the duration of the segment, with the first window centered at the start of the segment. For each measure, the values obtained for the middle three windows were then averaged to yield the mean value of either centroid or skewness for each token, on which the statistical analysis was performed.

As previously mentioned, we seek to understand how the realization of /v/ tokens compares relationally. Thus, in order to explicitly compare how closely paired [f] and [v] are in the three languages, all the segments had their measures relativized to [f], and speaker-specific differences were controlled for. More specifically, for each measure, for each speaker, the mean value of [f] was computed; this value was then subtracted from each value of that measure taken for that speaker. The result of this computation is that the mean value calculated for each measure for [f] is 0, and the measurement values for all other segments are expressed in terms of their distance to [f]. These relative measures were then averaged over speakers within a language. These measures will henceforth be referred to as the relative centroid and relative skewness, and it is understood that the measures are relativized to [f] in the manner described here.

For both measures, means were taken over all repetitions and speakers, and z-scores were calculated within each environment. Tokens were excluded if they had a z-score greater than 2.5 with respect to the mean value for that measure for a given segment, within a language.
3.3 Results

Figure 11.1 shows the normalized centroid for all three languages, where the relative centroids are plotted on the same axis; recall that since the centroids are relativized to [f], its mean relative centroid is 0.

First, notice that for all three languages, [z] has a similar relationship to [s]. The Greek sibilants are produced with a more palatal articulation (inferred by a lower spectral centroid) than in Serbian and Russian, but the key point is that the relationship between [z] and [s] is similar in all cases. This captures the fact that [s] and [z] comprise a voicing pair. Additionally, [m] in all languages stands on its own and does not pattern with the obstruents, as expected.

Figure 11.1 illustrates a dichotomy between Greek and Russian on the one hand, and Serbian on the other. The relative frication of [v] to [f] is the same in Greek and Russian, showing a similar relationship with respect to the distribution of high-frequency energy; this contrasts with Serbian, in which the relative frication of [v] is much lower. A two-way ANOVA (segment × language) on the relativized centroid values for [f, v, m, s, z] showed main effects of both segment [F = 380.98, p = 1.79944e-190] and language [F = 33.82, p = 7.0569e-015], as well as an interaction of segment and language [F = 33.75, p = 1.365373-0.46]. Post-hoc Tukey tests indicate that, as we expect from Figure 11.1, the spectral centroid values for [v] in Greek and Russian do not differ significantly from each other, but do differ from Serbian [v].

Post-hoc Tukey tests further show that within a language, the centroids for the sibilant pairs [s, z] do not differ significantly; thus, modulo the effect of voicing, the distribution of high-frequency energy in the sibilants is the same. Relevant to this study is the fact that post-hoc Tukey tests show that the spectral centroid for [v] does not differ significantly from that of [f] in Greek and Russian, but that the centroids for [v] and [f] do differ significantly in Serbian.

Figure 11.1  Relative centroid of [v] tokens in Greek, Russian, and Serbian.
Figure 11.2 shows the relative skewness for all three languages, and paints the same picture as above. For all three languages, the relationship between [s] and [z] is roughly the same, and indicates that these sounds comprise a voicing pair. Again, [v] stands in similar relation to [f] in Greek and Russian, but [v] has a very different relationship to [f] in Serbian.

A two-way ANOVA (segment $\times$ language) on the relativized skewness values for [f, v, m, s, z] showed main effects of both segment [$F = 275.14, p = 1.844e-153$] and language [$F = 11.28, p = 1.4569e-005$], as well as an interaction of segment and language [$F = 21.61, p = 3.4521e-030$]. Post-hoc Tukey tests indicate that the spectral skewness values for [v] in Greek and Russian do not differ significantly from each other, but do differ from Serbian [v].

These results indicate that in pre-vocalic, word-initial stressed position, tokens of Serbian [v] are produced with little energy in the high-frequency range, indicating a lack of frication, while tokens of Greek and Russian [v] are produced with frication in these environments. Crucially, there is no statistical difference between Greek and Russian based on the relationship of [v] to [f].

3.4 Discussion

Whether the segment phonologically transcribed as /v/ is realized as a fricative or as an approximant is, in articulatory terms, a question of labiodental aperture and the gradient between oral and atmospheric pressure. Acoustically, this is most likely to manifest in the presence of high-frequency energy, which results from the turbulence generated both at the place of stricture between the upper teeth and lower lip, and where the airstream is impeded downstream from the stricture at the upper lip (Stevens 1988). This was tested acoustically by calculating the spectral moments...
of centroid and skewness on the high-pass filtered signal (at 1500Hz) to ascertain the distribution of high-frequency energy.

The centroid and skewness of Greek /v/ differed significantly from those of Serbian /v/. Moreover, the normalization procedure showed that the noise distribution of the labiodental continuants in Greek parallels that of the sibilant continuants; in Serbian, the close relationship between the sibilants is not mirrored in the labiodentals, indicating that they do not form a voicing pair. We interpret these results to indicate that in Greek, /v/ is the voiced counterpart to /f/, both phonologically and phonetically, while in Serbian /f/ and /v/ do not comprise a voicing pair, either phonologically or phonetically. Russian /v/ exhibited values for both phonetic parameters that were not significantly different from Greek /v/.

These results suggest that the correlation between phonological status and phonetic realization of /v/ is relatively strong for Greek and Serbian, but rather weak for Russian. Specifically, Greek /v/ distributes and patterns as the voiced counterpart to /f/, and this is reflected in its phonetic realization: modulo the effect of voicing, the noise distribution of /v/ and /f/ are similar in Greek, supporting the classification of [v] as a voiced fricative obstruent. In contrast, Serbian /v/ distributes and patterns as a sonorant, and is not paired with /f/; this is reflected phonetically by its lack of frication. To better reflect both its phonological patterning and phonetic realization, Serbian /v/ would be better transcribed as /ʋ/. In Russian, though /v/ distributes as a sonorant, it patterns with both sonorants and obstruents. However, the results of the present study do not support a correlation between its phonological status and its phonetic realization; Russian /v/ patterned phonetically as Greek /v/, displaying energy in the high-frequency range indicative of a voiced fricative.

The environment selected for this study was chosen because it is considered amenable to subtle distinctions in consonant realizations: word-initial stressed syllables are prosodically strong, and consonants are most salient and most easily articulated in pre-vocalic position. Moreover, such an environment is most easily controlled cross-linguistically, allowing for a robust comparison of the acoustic realization of consonants in Greek, Russian, and Serbian. Nevertheless, further work is needed to assess the range of possible realizations of /v/ tokens. A future avenue of research is to extend the cross-linguistic approach taken here to studying the acoustic realization of /v/ in consonant clusters. Another important issue to further pursue is that of inventory structure. Greek and Serbian were selected to locally mimic the Russian inventory, containing /f/ and lacking /w/. An interesting extension of this study would be to explore whether the presence of /w/ or /v/ in addition to /v/, or the lack of /f/ has an effect on the phonetic realization of /v/.

### 4 Conclusion

This chapter has presented a case study of the segment typically transcribed as /v/ in an effort to shed light on the phonological and phonetic factors that contribute to sound classification. The motivation for this study derives from the ambiguous
The phonetics and Phonology of Segment Classification

phonological status of Russian /v/ with respect to the feature [sonorant], a problem well known in the phonological literature. This study situates the Russian case in a cross-linguistic setting, both phonologically and phonetically, comparing the distribution, patterning and phonetic realization of Russian /v/ with /v/ in languages where it is classified as an obstruent, as in Greek, and as a sonorant, as in Serbian.

The importance of the cross-linguistic approach can be seen by considering Padgett’s phonological analysis of Russian /v/. Padgett’s (2002) central claim is that the behavior of Russian /v/ derives from an interaction of its surface-based properties and a cue-based approach to phonology (Steriade 1997). Specifically, Russian /v/ is characterized by inherent phonetic properties intermediate between approximants and fricatives, and Padgett transcribes this sound as [γ]. The featural specifications of [v, y, v̞] given by Padgett (2002) are shown in Table 11.6 (vocoid) is simply the feature [consonantal] with values reversed. Padgett proposes the feature [wide] as a non-contrastive feature that refers to constriction degree. Obstruents, having a narrow constriction degree, are specified as [‐wide], while sonorants are [+wide]; Russian [y] is specified as [+sonorant] and [‐wide], thereby featurally encoding both its obstruent and sonorant characteristics.

Such a segment is inherently unstable due to the aerodynamic constraints of maintaining both frication and voicing, and so [γ] can only be realized (and perceived) as such in pre-sonorant position (in fact, pre-vocalic position ought to be the most amenable to such a realization). Such an analysis relies on there being a strong correlation between the phonological status of /v/ and its phonetic realization, but as we have seen, the correlation is somewhat weaker. In particular, despite the differences in the phonological status of /v/ between Russian and Greek, the degree of frication was not found to differ significantly between Russian [v] tokens and Greek [v] tokens. The results of this study suggest that the phonological identity of a segment cannot be wholly determined by its phonetic properties, and in particular that a cue-based approach does not accurately predict phonological classification, as we would expect either Greek /v/ to pattern as in Russian, or Russian /v/ to pattern as in Greek, which is not the case.

The results of this study highlight the need for tightly controlled, cross-linguistic phonological and acoustic studies in trying to establish how sounds are classified, incorporating considerations such as inventory structure, distribution, phonological

<table>
<thead>
<tr>
<th></th>
<th>Stop</th>
<th>Fricative</th>
<th>Narrow Approximants</th>
<th>Wide Approximants</th>
<th>Glide</th>
<th>Nasal</th>
</tr>
</thead>
<tbody>
<tr>
<td>[continuant]</td>
<td>b</td>
<td>v</td>
<td>y</td>
<td>v</td>
<td>w</td>
<td>m</td>
</tr>
<tr>
<td>[approximant]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[wide]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[vocoid]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[sonorant]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
patterning and phonetic realization. Only through such thorough investigations can we understand whether there exists a correlation between the phonological status of a segment and the phonetic parameters characterizing the realization of tokens of that segment.

Acknowledgments

Special thanks go to Alexei Kotchetov at the University of Toronto for allowing me to use the Phonetics Lab there, and the linguists at Cornell University for helpful discussion.

References

Padgett, J. (2002). Russian voicing assimilation, final devoicing, and the problem of [v] (or, The mouse that squeaked). MS.
Part III

Case Studies
The Perception of Vowel Quality and Quantity by Turkish Learners of German as a Foreign Language

Katharina Nimz

1 Introduction

Of all aspects of second language phonetics and phonology it is probably the segment that has attracted most attention (Archibald and Young-Scholten 2003; but see Altmann and Kabak 2011 for a recent overview on segmental as well as supra-segmental L2 phonology). Both early works on L2 phonology, such as Lado (1957) in his famous contrastive analysis hypothesis, as well as the most prominent and recent models such as Flege’s (1995) Speech Learning Model or Best and Tyler’s (2007) Perceptual Assimilation Model for L2 perception, focus on the acquisition of the L2 segment. The main aim of this chapter is to report on experimental research into L2 vowel perception by foreign language learners which can shed some light on what are important cues when learning novel units of speech, in this case, German vowels. In the following, I will give a short overview of related work in the field of L2 vowel (length) perception, briefly compare the German and Turkish vowel systems – supported by a small set of production data which were collected during the study as well – and present a perception experiment that was carried out with Turkish German as a Foreign-Language (GFL) learners on the perception of the German point vowels. The stimuli in the discrimination experiment were manipulated in a way that made it possible to distinguish between the dimensions quality (spectral cues) and quantity (durational cues), both of which are used differently in German and Turkish.
2 L2 vowel perception

The idea that our native phonological system will influence the perception of foreign sounds is not new. Trubetzkoy (1939/1969) in his seminal work on the principles of phonology had already written about L2 sound perception and claimed that our native phonological system functions as a “sieve” in that it filters out those properties of the speech signal that are not relevant to the native phonological system, consequently leading to mistakes in the perception (and the production) of foreign sounds. Experimental work on L2 vowel perception corroborates this assumption: Learners of an L2 have been found to differ from native (L1) speakers of the language in their use of acoustic cues and consequently differ in their perception from native speakers (Bohn and Flege 1990; Bohn 1995; Flege 1995; Flege et al. 1997; Flege et al. 1999; Levy and Strange 2008; or Darcy and Krüger 2012, among many others).

In his famous Speech Learning Model (SLM), Flege (1995) accounts for findings which have shown that not all foreign sounds are necessarily difficult for L2 learners to acquire. An important concept in this context is that it is not the different sounds (as for example postulated in the contrastive analysis hypothesis by Lado 1957, see below), which are difficult for L2 learners, but the similar ones. Flege assumes that the mechanisms used in L1 phonetic category formation are intact over the life span and can be applied to L2 learning, that is, it is possible that learners can build new L2 phonetic categories and, with that, perceive and produce L2 segments like a native speaker. This new category formation becomes more likely the younger the learners are and the more perceptually different the closest L1 and L2 sounds are. If the sounds are too similar, new category formation is blocked, as so called equivalence classification will take place, that is, new sounds are falsely processed as L1 sounds. Differentiation between different types of cues in vowel perception, for example spectral and durational cues, is not specifically addressed in the model, although it is discussed in related articles (Bohn and Flege 1990; Bohn 1995).

In GFL research, it is still common to rely on contrastive analyses between L1 and L2 sound systems in order to predict possible interferences in the sound productions of foreign language learners (Redecker and Grzeszczakowska-Pawlikowska 2009: 116). In contrast to the experimental studies cited earlier, the main focus of this line of research is on speech production, probably due to its focus on practical applications in foreign language teaching. The idea behind contrastive analysis (Lado 1957) is that sounds which do not occur in the L1 will be difficult to learn in the L2. Accordingly, GFL researchers have predicted that both German vowel quality and quantity are difficult features for Turkish GFL learners to acquire, because both dimensions differ in the two languages (Neumann 1981; Slembek 1995; Rolffs 2005).

2.1 L2 vowel length perception

Most studies on L2 vowel perception have not directly addressed the issue of vowel length and how this cue might be used in the acquisition of new sound
categories, with notable exceptions such as Bohn (1995), McAllister et al. (2002) or Cebrian (2006). Most of the experimental work in L2 phonetics and phonology was done on (North American) English as an L2, a language which is said to use duration as a secondary cue to vowel perception (Bohn 1995; Hillenbrand, Clark, and Houde 2000). Based on his results from studies with German, Spanish, and Mandarin native speakers identifying synthetic English vowels, Bohn (1995) came to the conclusion that duration cues in vowel perception are easy to access whether or not listeners have had experience with them in their native language (Desensitization Hypothesis). In his studies, the participants had to identify English vowels on a bet to bat (and beat to bit) continuum as bet or bat (beat or bit), which were manipulated in duration and spectral features in equal steps. German native speakers were found to rely much more on the durational differences when identifying stimuli as bet or bat than the native English speakers; interestingly, the same was true for the Spanish and Mandarin participants for the beat to bit continuum, who, in contrast to German speakers, do not use duration as a distinctive feature in their native languages. Cebrian (2006) conducted similar studies with different groups of L1 Catalan learners of L2 Canadian English. His results supported Bohn’s Desensitization Hypothesis in that the Catalan speakers relied on duration as the main cue to the English vowel contrast despite not having experience with duration in their native language.

McAllister et al. (2002) investigated the perception (and production) of Swedish quantity distinctions by Estonian, English, and Spanish L2 learners of Swedish. The three phonologies of the learners’ native languages display different degrees of overall prominence of the duration feature: Estonian makes use of phonological vowel and consonantal length contrasts, while in English, length is considered a secondary cue in vowel perception, and Spanish does not have any phonological length contrasts at all (hence: Estonian > English > Spanish). It was expected that Estonian speakers could perceive the difference between long and short Swedish vowels better than the English or Spanish speakers, and that the English speakers would be better than the Spanish. In contrast to Bohn (1995) and Cebrian (2006), the identification task of McAllister et al. involved real Swedish words and non-words, which were created by replacing the long vowels (and following short consonants) by short vowels (and following long consonants). Participants then had to judge whether the stimuli were correct or incorrect instances of the respective words. Since the Estonian speakers were better at identifying the test words correctly than the English and Spanish speakers, and since the English were better than the Spanish (for some of the vowel pairs), their results led McAllister et al. to formulate the Feature Hypothesis, which states that L2 features, such as duration, not used to signal phonological contrast in the L1 will be difficult to perceive for L2 learners, which contrasts with what is implied by the Desensitization Hypothesis, namely that duration cues are easy to access despite the native language background. These different hypotheses will become relevant again in section 4.
3 The German and Turkish vowel systems

Though there has been rigorous critique as to how gainful a contrastive analysis of sound systems (and their phonetic symbols) may be for the formulation of hypotheses concerning L2 speech (Bohn 2002), in the following I will give a very brief contrastive overview of the two vowel systems that are of interest in this chapter (see Figure 12.1).

What becomes immediately apparent when looking at the two IPA vowel charts is that German has about twice as many vowels as Turkish. This is due to the fact that in stressed syllables German makes a phonological contrast between short/lax and long/tense vowels. In the IPA chart, the tense vowels are not marked for length (with the exception of /ɛ:/ and /a:/), because it can be inferred from the tenseness of the vowels. Of interest for the present study are the German point vowel pairs /i:/-/ɪ/, /a/-/a:/, and /u:/-/ʊ/. Similar vowels exist in the Turkish vowel system (circled in Figure 12.1, right chart). As can be deduced from the left chart, the German low vowels /a/ and /a:/ are assumed to be identical in vowel quality, as they are depicted with the same tongue position; both tense /i:/ and tense /u:/ however differ from lax /ɪ/ and /ʊ/ in that their tongue position is more peripheral, that is, /i/ is articulated with a tongue position that is much higher and more forward than for /ɪ/, while /u:/ is articulated further back and higher than /ʊ/, though the difference does not seem to be as extreme as for the “i-pair” (I will, for reasons of convenience, refer to the point vowel pairs as “i-pair,” “a-pair,” and “u-pair”).

Turkish /i/, /a/, and /u/ are represented with the tense vowel symbols; however, it is unclear if the quality of the Turkish vowels really is identical to that of the German tense vowels. The IPA charts are articulatory approximations and not meant for precise acoustic comparisons. Truly comparable acoustic data, that is, data that were collected for the German and Turkish group in a comparable fashion (speakers’ sex, number of speakers/items, experimental set-up, etc.), did not exist prior to the study, which is why such data were collected as well (see Figure 12.2).

According to this exploratory dataset it seems that Turkish /i/ is almost identical to German /i/, in contrast to German /ɪ/, which seems relatively distant from the

![Figure 12.1](image_url) German (left) (Kohler 1999: 86) and Turkish (right) (Zimmer and Orgun 1999: 154) vowel systems as depicted in the Handbook of the IPA. Vowels of interest to the study have been circled.
Turkish vowel in the F1-F2 acoustic space. Turkish /u/, however, seems to be equally similar to German /u/ as it is to German /ʊ/. Turkish /a/ is equally far away from German /a/ and /aː/.4

In Turkish there exists no phonological opposition between short/lax and long/tense vowels. It is a vowel system that is generally described as a very symmetrical vowel system of eight vowels which systematically differ in the three dimensions high/low, front/back, and rounded/unrounded lips (Lewis 2000; Moser-Weithmann 2001, or Göksel and Kerslake 2005). While Turkish is phonologically described as a language without long vowels (Kabak 2004), there are two occasions in which long vowels do occur: first, the Turkish language uses a considerable amount of borrowed words from Arabic, a language, which has phonological vowel length contrast. Though Turkish phonology does not make use of a length contrast itself, the long vowels of the borrowed words are still preserved in Turkish pronunciation (e.g., saat [saat], English"hour"). Second, so called secondary long vowels can be found as a result of compensatory lengthening (Kabak 2007): the Turkish voiced velar fricative /ɣ/ (in orthography represented as ğ (yumuşak ğ)) is not pronounced in Turkish, resulting in compensatory lengthening of the preceding vowel. Furthermore, it should be mentioned that Turkish has geminate consonants, for example eli (English “his hand”) but elli (English “fifty”) (Lewis 2000). Still, the vowel length in Turkish is not distinctive as it is in German, which is why we wanted to shed light on the question of whether Turkish learners of GFL have problems with vowel length in German or not.

### 4 Research questions and method

The main research question was whether it is vowel quantity and/or quality that is problematic for Turkish learners of German as a foreign language to perceive. On the basis of traditional contrastive analyses, GFL researchers have made the claim that it is both dimensions that pose a problem for learners, as Turkish vowels differ

---

**Figure 12.2** Acoustic data of German and Turkish point vowels. Mean formant values for three different one-syllable German/Turkish words (three repetitions). Four male speakers per language group.
both in their vowel qualities and their use of vowel quantity from German vowels. In the field of experimental L2 phonetics and phonology, however, there exist two contradicting hypotheses that are specifically concerned with the perception of vowel quantity: While Bohn’s Desensitization Hypothesis would predict that Turkish learners would not have problems perceiving quantity differences in German vowels, as duration cues are supposedly easy to access for any group of native speakers, McAllister et al.’s Feature Hypothesis would predict the opposite, as vowel length differences are not employed in the phonology of Turkish as they are in German. Accordingly, this study further set out to shed light on these opposing predictions. As far as vowel quality is concerned, it was expected that not all vowel quality differences would be equally difficult to differentiate for the Turkish learners, as the Turkish vowels are not equally similar/different from the German vowels under investigation. However, since comparable acoustic data did not exist prior to this study, the predictions for the quality dimension could not be more specific.

Hypotheses were tested using an AX discrimination task.

Since a discrimination task using naturally occurring German vowels alone would not have allowed for differentiation between the dimensions quality and quantity, it was necessary to manipulate the vowels so that stimuli would either be different in their quality or in their length. This was done in a similar way to a design used by Sendlmeyer (1981), who was interested in the perception of long/tense and short/lax vowel oppositions by German native speakers. In his study he had lengthened the short vowels of minimal pair words to the length of their long counterparts and had shortened the long vowels to the length of their short counterparts. He then had native speakers judge the manipulated words as to which real word they heard. What he found was that for the German high vowels, quality seemed to be the primary cue, while for the “a-pair” and some of the mid-vowels, quantity is the decisive feature (see also Weiss 1974).

### 4.1 Stimuli

Because the performance of Turkish GFL learners compared to German native speakers was of interest, nonsense words were chosen as stimuli, namely vowels in the bilabial consonantal frame [b_p], as it would have been difficult to control for the factor of familiarity with real test words. The test items were recorded in the frame *Ih hab einen [...] gesehen* (“I have seen a [...]”) and were produced by a female native German speaker from the west of Germany. Of five productions of each lax and tense point vowel (plus the vowels [ɛ, ɛ, ɔ, ɔ] for a test block before the actual experiment), those productions were chosen for further analysis which were closest to the speaker’s mean values of the first and second formants of the respective vowel (measured at mid-point). These items were then used for further manipulation.

Since the test items were chosen based on their formant values, their lengths were adjusted so they would match the calculated means of the respective short or long
The Perception of Vowel Quality and Quantity by Turkish Learners

This was done with the help of PRAAT (Boersma and Weenink 2008), by cutting out or replicating whole periods from the middle of the vowels. The experiment included three experimental conditions: “proto,” “length,” and “quality.” The test items of the “proto” condition were the ones just described, that is, the items that best matched the mean formant values of the speaker plus minor length corrections of one or two periods. In the other two conditions items had to be shortened or lengthened much more, as in the “length” condition the two contrasted items were of different lengths but of the same quality, and in the “quality” condition the two items were of different qualities but of the same length. As in Sendlmeier (1981), this was achieved by lengthening the short vowels to the length of the long vowels and vice versa. Furthermore, the consonantal environments were adjusted in that all voicing action of the initial consonants was set to zero and all bursts were the same for the contrasted items by cutting and pasting the consonantal sequences at the respective place of the corresponding counterpart. Accordingly, the following experimental items had to be discriminated in the experiment (here, the “u-pairs” as illustration):

(1) a. Condition ‘proto’ (prototypical short/lax [ʊ] versus prototypical tense/long [uː] → [bʊp] versus [buːp])
   b. Condition ‘length’ (prototypical long/tense [uː] versus shortened long/tense [u] → [buːp] versus [bup])
   c. Condition ‘quality’ (prototypical long/tense [uː] versus lengthened short/lax [ʊː] → [buːp] versus [bʊːp])

Furthermore, a control pair was included in the experiment ([bap] versus [bɪp]), to ensure that the participants really understood the task and did not just answer randomly, plus six filler pairs that were the same (by pairing each “proto” vowel ([iː], [ɪ], [a], [aː], [uː], [ʊ]) with itself). Every pair was played to the participants five times, which in all yielded 80 pairs to be judged: 5 × (3 vowels × 3 experimental conditions) + 5 × 1 control + 5 × 6 filler pairs. The stimuli within a pair were concatenated without any inter-stimulus interval (ISI); between the pairs, participants had two seconds to give a same/different response on a corresponding spreadsheet. Stimulus pairs were presented in blocks of eight with a 10-second pause between each block. In all, the experiment lasted about five minutes, including a small test block with different vowel pairs in which feedback was given to ensure that the participants understood the task.

4.2 Participants

The responses of 20 Turkish participants (Ø age 17.5, SD = .5) and 20 German participants as control group (Ø age 17.9, SD = .7) were analyzed for the experiment. The Turkish learners were students at a German high school in Istanbul, where they had received at least three years of intensive German as a Foreign-Language classes.
In the first year, they received about 20 hours of teaching by German native speakers and in the following years about 10 hours per week, while other subjects such as biology or mathematics were also taught in German. They reported mainly speaking Turkish at home and with their peers and none of them had spent a significant amount of time in a German-speaking country, which makes them rather prototypical GFL learners (as opposed to GSL learners). The German native speakers were high school students in the west of Germany, where Standard German, the variety taught to the Turkish learners, is spoken.

5 Results

Judgments differed greatly for each vowel group which is why the scores of the German natives and Turkish GFL learners had to be analyzed separately for the experimental conditions for the “a,” “i,” and “u” group. Also, since the data were not normally distributed, the non-parametric Mann-Whitney U-test was used to look for significant group differences between the German and Turkish participants. Since multiple tests (nine comparisons) were conducted, the significance level was adjusted according to the Bonferroni correction to $\alpha = .006$. Figure 12.3 shows the percentages of correct responses for the three different vowel groups.

5.1 “a”-group

There were no significant group differences between the German natives and Turkish GFL for any of the conditions within the “a”-group. Though the Turkish participants heard the difference between a long [a:] versus its shortened version...
The Perception of Vowel Quality and Quantity by Turkish Learners

(261) (length condition) on average in only 87% of the cases, while the German natives could hear this difference in almost all pairs, this difference was not significant with the adjusted α (p = .03, U = 147.5). At first sight, it seems striking that neither group could hear the differences in the quality condition. However, this result is easily explained when keeping in mind that the unmanipulated German short and long low vowels are assumed to only differ in length, hence, if this parameter is manipulated in the way that both vowels have the same length, they should sound exactly the same – which they do for both the German and the Turkish participants. Since the focus was on specific problems of the Turkish learners as compared to the German natives, however, this specific result is not of particular interest.

5.2 “i”-group

As in the “a”-group, vowels of the proto condition in the “i”-group could be differentiated by the Turkish just as well as by the German natives. Apparently, long/tense /i:/ is different enough from short/lax /i/ also for learners whose language does not differentiate between lax/tense or short/long. From the results of the other two conditions in this vowel group, this very good discrimination (98%) seems to be due to both the differences in length and quality between the two vowels, as the Turkish participants heard the spectral or duration differences equally well (quality condition at 93%, length condition at 92%), and so did not significantly differ from the German natives. The German natives were slightly better at discriminating the quality difference than the duration differences for the i-pairs. Though again not part of the main research question, it is interesting by itself as this goes in hand with the assumption that for the high vowels, quality is the primary cue for German native speakers (Sendlmeier 1981).

5.3 “u”-group

The most interesting pattern as regards the research question can be found for the u-pairs. There was a significant difference between the German and Turkish participants in the quality condition (p < 0.001, U = 42.5): Turkish participants misjudged the pairs in which the vowels only differed in their quality 44% of the time, while German native speakers could hear this difference 92% of the time. Where differences existed only in length, the participants could hear the differences equally well – or rather equally badly – with about 69% in the German group and 64% in the Turkish group (length condition). Most interestingly, a significant difference between the groups existed also in the proto condition (p = .003, U = 120), the condition which is met in real world situations when faced with German minimal pairs such as spuken [ʃpuːkən] (“to haunt”) versus spucken [ʃpɔkən] (“to spit”) or Buße [busə] (“repentance”) versus Busse [bʊsə] (“buses”).
5.4 Discussion

The research question was whether it is quantity and/or quality that is problematic for Turkish GFL when hearing German vowels. It became apparent from the data that it is not possible to draw a uniform picture for all German long/tense and short/lax vowel pairs. Turkish learners could distinguish well between prototypical \([a:] / [a]\) and \([i:] / [i]\); however, they had problems distinguishing \([u:] / [u]\). Why is that? When looking at the data within the manipulated “length” and “quality” conditions, it seems obvious that it must be due to the fact that the spectral differences between \([u]\) and \([u]\) are difficult for Turkish learners to hear. Though they did not hear the differences in all the cases of the length condition for the “u”-pairs, either, they did not differ in that respect from the German natives. In general, it seems that length is less important for distinguishing the u-pair in German, which is reflected in the fact that the ratio for the long and short a- and i-pairs is bigger than for the u-pair. Since the Turkish GFL learners did not differ significantly from the German natives in either of the length conditions, that is, in those conditions where vowels differed only in length, the data suggest that Turkish GFL learners do not have specific problems hearing the difference in length between different types of German vowel contrasts, at least not more so than German native speakers.

There are two possible explanations for this. Either this finding supports assumptions made by Bohn’s Desensitization Hypothesis, or, the less exciting one, Turkish learners can simply hear the length difference because they have experience with it on a phonetic (as opposed to phonological) level in their native language. Bohn (1995) stated that duration cues are easy to access whether or not a listener has had experience with them in her native language or not, which could explain why Turkish learners could differentiate all vowels in the length condition. Though length is not used as a vowel feature in the phonology of Turkish, one could also argue that the learners’ experience with vowel length in loanwords and through secondary lengthening of their native vowels – or even the experience with the consonantal length feature – is prominent enough to exert a kind of positive transfer, which is why the Feature Hypothesis by McAllister et al. (2002), or assumptions made by GFL researchers based on contrastive analyses, cannot be ruled out.

The different results for the quality conditions for the different vowels can be explained when taking into account the short acoustic comparisons made in section 4 and how they could relate to assumptions made by Flege’s SLM. The vowel qualities that were problematic for the learners to differentiate were \([u:]\) versus \([u]\). Since Turkish /u/ seems to be equally far away from the lax and tense German counterparts, one could claim that they are equally similar, and so are both subject to “equivalence classification” with the Turkish vowels, that is, they cannot be well differentiated from Turkish /u/ or between each other. In the case of /i:/ and /i/, it could be argued that German /i/ is sufficiently different from Turkish /i/ and can therefore be well distinguished from German /i(:)/, which is almost identical to Turkish /i/. Certainly, this kind of explanation is only tentative, and the acoustic data could not be used to precisely predict the obtained results, as both sets of data were collected simultaneously.
6 Conclusion

Contrastive analyses of the phoneme systems of Turkish and German have led various GFL researchers to the conclusion that Turkish learners of German as a Foreign Language have problems with the short/lax versus long/tense distinction in German vowels. Our results suggest that it is not necessarily the length of the German vowels but rather the different qualities, that is, lax and tense, that are specifically problematic for the learners. However, the generalizability of our results is necessarily limited, as the discrimination task was of a very basic kind that is rarely met in real life communication situations. Still, it is striking that despite the ease of the task, Turkish learners only heard the difference between manipulated lax, long [ʊː] and prototypical tense [uːː] in less than 50% of all instances, and further, most importantly, heard the difference between the prototypical u-pair items significantly less often than the German natives. Since our significant results are limited to the “u”-group, it would be interesting to further investigate how participants would perform under more difficult listening situations with other vowel pairs (i.e., background noise, larger ISI, different speakers, more vowel contrasts, surrounding context, etc.). Furthermore, it would be informative to conduct similar experiments with native speakers of a language which does not exhibit phenomena such as secondary lengthening of vowels (see for example Nimz Forthcoming).

Despite certain limitations of the study, it is possible to apply the results in GFL teaching: though the exact reasons are unknown, Turkish learners seem to be able to exploit the quantity dimension in German vowels just as well as German native speakers. Teachers of German as a Foreign Language might therefore be more target-oriented if concentrating on the different vowel qualities (with the exception of /a/ and /aː/) rather than the differences in quantity. The study of the L2 segment is therefore not only of theoretical interest but finds its relevant and practical application in the foreign language classroom.

Acknowledgments

I would like to thank Bernd Pompino-Marschall, Marzena Żygis, Sabine Zerbian, and Ghada Khattab, as well as the delegates at the ICPhS XVII in Hong Kong and the CUNY Phonology Forum Conference 2012 in New York for their support and insightful feedback. I further thank the schools, teachers, and participants who made this research possible.

Notes

1 German as a Foreign Language is an interdisciplinary field of research with ties to areas as diverse as pedagogy, social sciences, literary and cultural studies, linguistics, and second language acquisition (Barkowski and Krumm 2010). In the German language and
L2 research traditions, there is a rather strict division between German as a Foreign Language (GFL), with a focus on L2 acquisition in a classroom setting, and German as a Second Language (GSL), with focus on L2 language acquisition of immigrants in a naturalistic setting. The “cover term” SLA is not used as frequently in German as it is in English.

In Swedish stressed syllables, there is a complementary relationship between the duration of the vowel and the consonant: a long vowel is followed by a short consonant, and a short vowel is followed by a long consonant (or a cluster).

Of course, it could also be the other way around, as tenseness (with the exception of /a:/ and /ɛ:/) could be inferred from the length of the vowels as well.

The terms “similar” and “different” play an important role in the framework of the above mentioned SLM, yet are problematic terms due to their relative nature. Though distances in an F1-F2 acoustic space have been used to refer to auditory distances (Flege et al. 1994), this assumption needs to be viewed with caution. However, recent research by Escudero et al. (2012) suggests that the concept of similarity is after all closely related to the detailed acoustic properties of sounds.

Even though the Turkish participants were advanced learners of GFL (at least three years of intensive GFL lessons), it was very unlikely that they would know all the words constituting possible minimal pairs in German.

Calculated mean values of produced items of the speaker: /a:/ (166 ms), /a/ (73 ms), /i:/ (125 ms), /ɪ/ (55 ms), /u:/ (119 ms), /ʊ/ (62 ms).

Long/short ratio for both the i- and a-pairs: 2.3, for u-pair: 2.2 (see previous footnote for mean values in ms). Antoniadis and Strube (1984) also found smaller ratios for the u-pair than for the i- and a-pair.

This, however, seems to be very unlikely. Flege (1995: 267) discusses findings of Flege and Port (1981) in the light of “free feature recombination” for the voiceless feature in stop consonants by Arabic speakers of L2 English. They found that it was not possible for the L2 learners to transfer the voiceless feature of their /t/ and /k/ to /p/. Since it does not even seem to be possible to transfer features within one natural class, it is highly unlikely that it is possible to transfer a feature used for constants to vowels.

In the terms of Best’s (1995) Perceptual Assimilation Model (PAM), this situation would be called “Single-Category Assimilation,” which predicts poor discrimination as well.

References


Nimz, K. (Forthcoming). Differences in the perception of German vowels: the case of Turkish and Polish learners of German as a Foreign Language. Proceedings of LingBaW 2013, Lublin.
13

Compensatory Lengthening in Hungarian VnC Sequences

Phonetic or Phonological?

Mária Gósy and Robert M. Vago

1 Introduction

Compensatory Lengthening (CL) is a widespread phenomenon in the languages of the world. In its most typical and simplest configuration, a coda consonant drops out, and its loss is compensated for by lengthening the preceding vowel. For an in-depth exposition of the salient facts and issues over a typologically diverse set of languages, see Wetzels and Sezer (1986), Hayes (1989), and Kavitskaya (2002), among others.

CL, by its intrinsic nature, provides strong support for the reality of segments. Let us assume, rather uncontroversially, that the phonology-to-phonetics channel postulates two levels of representation (recognizing that some models allow for additional intermediate levels): abstract vs. concrete, or underlying vs. surface, or input vs. output, or phonological vs. phonetic, and so on. Staying with the input vs. output nomenclature, we may call an overt segment derived if its output form does not correspond to its input form. The disparity can be complete, as in the case of an epenthetic segment, or partial, as in the case of change in feature composition (e.g., assimilation) or prosodic affiliation. A different group of segments, call them ghost segments, have input forms that do not have corresponding output forms. Again, this divergence can be complete, as in the case of deletion that leaves no trace in the output (dropping lock, stock, and barrel), or partial, as in the case of deletion that leaves some featural or prosodic effect in the output (typically, on a neighboring segment).

Within the context of the above typology, CL evidences both partial-derived segments and partial-ghost segments. Consider the quintessential CL case: VC in the input corresponds to Vː (= long vowel) in the output. On the one hand, C functions as a partial-ghost segment in that even though it does not appear in...
the output, it has lent one of its abstract structural units (mora or skeletal tier – depending on one’s theoretical framework) to the preceding V, in effect inducing (phonological) length (Vː). And on the other hand, the resultant output unit Vː is a partial-derived segment in that its input form (V) has acquired the prosodic property of length.

In this chapter we aim to establish the facts of a rather complex CL process in Hungarian, based on acoustic phonetic and statistical analyses of data collected in an experimental study. The results, we argue, suggest that this particular CL finds its motivation not in phonological structure, but rather in articulatory phonetic implementation, creating non-contrastive partial-derived segments.

2 Data

Hungarian evidences a number of variable or optional CL processes, whereby in casual speech a medial sonorant consonant ([l r j n]) may be deleted in the coda position of a syllable and, as a result, the preceding vowel is lengthened (Vago 1998; Siptár and Törkenczy 2000, among others). Here, we will be concerned with only those cases where CL is induced by the loss of /n/ in VnC sequences, as in the following examples:

Each of the above phonetic representations has a variant without the effects of CL, namely containing a short vowel followed by the nasal consonant, whose place of articulation agrees with that of the following consonant. Thus for example, vonz-ő may be pronounced in casual speech as either [võːzoː] or [vonzoː].

A number of observations are noteworthy about the data: first, the vowel before underlying /n/ becomes nasalized – even though the nasal consonant is not pronounced. Second, in all cases, underlying /n/ is followed by another consonant, which may be: (a) tautomorphic with /n/ (e.g., impotens); (b) the initial segment of a following suffix (e.g., latin-ra); or (c) the initial segment of a following word (e.g., ötven zsák). Third, the consonant following /n/ is a continuant (obstruent fricative, lateral approximant, or central approximant). In point of fact, if /n/ is followed by a noncontinuant (stop or affricate), CL does not obtain. Note the following examples:
(2) negyven perc ‘forty minutes’ [nɛɟvɛmpɛɾʦ] (*[nɛɟvɛːpɛɾʦ])
nagyön buta ‘very stupid’ [nɒŋɔmbutn] (*[nɒŋɔːbutn])
tinta ‘ink’ [tintn] (*[tɪttn])
sündisznó porcupine’ [ʃyndisnoː] (*[ʃyːdɪsnoː])
ötven tyúk ‘50 hens’ [ɔtvɛnuːk] (*[ɔtvɛːcuːk])
igen gyakran ‘rather often’ [ɪɡɛɲɔkɾɔn] (*[ɪɡɛːɲɔkɾɔn])
bank ‘bank’ [bɒŋk] (*[bɔːk])
barlang ‘cave’ [bɔrlɒŋɡ] (*[bɔrlɔːɡ])
kuncog ‘chuckle’ [kʊŋɔɡ] (*[kʊːʊɡ])
narancs ‘orange’ [nɔɾɔŋʧ] (*[nɔɾɔːʧ])
findzsa ‘tea-cup’ [fɪndʒsa] (*[fɪːʤsa])

3 Subjects, material, method

Eight native Hungarian speakers (four women and four men) with no known speech or hearing defects read isolated words in a sound-proofed chamber. Their ages ranged from 24 to 32. The word lists consisted of Hungarian words and phrases that contain the dento-alveolar nasal /n/ followed by either a fricative or a stop / affricate consonant; these represent the possible and impossible CL contexts, respectively. There were four vowels used in the material, two front (i [i], e [ɛ]) and two back (o [o], a [ɒ]); cf. Table 13.1.

The words were recorded and digitalized up to 44,000 Hz. Acoustic phonetic analysis was carried out by Praat software (Boersma and Weenink 2004). The consonant /n/, if present, the previous vowel, and the following consonant were defined in each word for each speaker. The duration of the nasal consonant and the preceding vowel, as well as the first, second, and third formants of the vowel were measured. The duration of the vowel was measured as the interval between the onset of the second formant and the onset of the nasal formant. Nasal duration was measured to the onset of the obstruent consonant and to the onset of the second formant of the approximant. The formant values were measured at the midpoint of total vowel

<table>
<thead>
<tr>
<th>Consonant following /n/</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>sz [s]</td>
<td>olyan szép</td>
</tr>
<tr>
<td>z [z]</td>
<td>nagyon zavar</td>
</tr>
<tr>
<td>s [ʃ]</td>
<td>latin-ság</td>
</tr>
<tr>
<td>zs [ʒ]</td>
<td>nagyon zsibbad</td>
</tr>
<tr>
<td>t [t]</td>
<td>latin-t</td>
</tr>
<tr>
<td>d [d]</td>
<td>olyan dagadt</td>
</tr>
<tr>
<td>c [ts]</td>
<td>olyan cukros</td>
</tr>
<tr>
<td>cs [ʧ]</td>
<td>istencsapás</td>
</tr>
</tbody>
</table>
duration. The F2 and F3 midpoints were estimated using visual inspection of wide-
band spectrograms, narrowband fast Fourier transforms (FFT), and auditory per-
ception. When /n/ dropped out and the preceding vowel became nasalized, the
duration of the /n/ “realization” was set to 10 milliseconds in all cases, in order to
avoid the problem of the different duration values of the vowels in question. (The 10
milliseconds is less than the shortest nasal transition in our material.)

The statistical evaluation of the data was carried out using ANOVA and Pearson’s
correlation test by SPSS 12.0.1 for Windows software package and regression anal-
ysis. In all cases, the confidence level was set at the conventional 95%.

4 Results

We will tabulate the two components of CL (loss of /n/ and vowel lengthening) in
our experiment individually.

4.1 /n/ loss

The principal question to raise is: why does /n/ drop out – or change its acoustic
structure – before continuants but not before noncontinuants? The explanation lies
in the basic difference between the articulation of /n/ plus continuant clusters vs. /n/
plus noncontinuant clusters. In general, /n/ takes on the place of articulation of a
following consonant, forming a homorganic cluster (Siptár and Törkenczy 2000:
207ff.). This is seen most clearly in cases where /n/ is followed by a noncontinuant
consonant. Thus, /n/ becomes [m] before [p], [b], or [m], where the entire cluster is
formed by a single closure with the lips; in the dento-alveolar cluster [n] plus [t], [d],
[n], [ʦ], or [ʣ], a single closure is made by the tongue at the teeth or the alveolus; in
the alveo-palatal cluster [n] plus [ʧ] or [ʤ] a single closure is made with the tongue
in the alveo-palatal area; in the palatal cluster [n] (= /n/) plus [c] or [j] a single
closure is made in the palatal area; and in the velar cluster [ŋ] (= /n/) plus [k] or [g]
a single closure is made in the velar area.

In all of these cases, the closure gesture is shared by the nasal consonant and the
following stop or affricate. The two closures – the one that belongs to the nasal and
the one that belongs to the stop or affricate – follow each other in time, but meld into
one gestural movement, since normally the closure of the nasal is not released before
the homorganic closure of the following noncontinuant. This mode of articulation
enhances the retention of /n/ so as not to eliminate the closure of the following non-
continuant consonant, motivated by maintaining the stability of the quintessential
consonantal articulation, namely the blockage of airflow.

In contrast, the dynamics of the articulation of /n/ in the case of a following con-
tinuant consonant is quite different (except in careful pronunciation, where /n/
undergoes regressive place assimilation). Anticipating lack of closure in the articu-
lation of the following consonant, the closure gesture of /n/ is eliminated in the oral
cavity, allowing the air to flow through both the oral and nasal cavities.\(^5\) (The nasal consonant is weak in this context: if pronounced, its duration is about half of the nasal that occurs before a noncontinuant. See Figure 13.3 below.) Most typically, the consequence is that the nasal consonant is realized as a transitional phase: the preceding vowel is nasalized to various degrees (see discussion below).

The two cases of VnC sequences have different acoustic consequences. Before noncontinuants, there is a well-defined nasal consonant, including the closure gesture, and the preceding vowel is short and non-nasal (with respect to Note 3); see Figure 13.1.

It turns out that before continuants the facts are much more complex than the phonological descriptions would indicate: the realization of /n/ falls into four distinct patterns. In the first one, the nasal consonant is missing and the preceding vowel is somewhat long and slightly nasalized; see Figure 13.2.

In the second case, the vowel is long and heavily nasalized, and there is no trace of the nasal consonant; see Figure 13.3.

In the third case, typically found in careful or deliberate speech, the nasal consonant is present, meaning the closure is actualized, and the vowel is short and non-nasal; see Figure 13.4. In other words, CL does not obtain.

In the fourth case, there is no trace of the nasal consonant – the /n/ drops out and the preceding vowel remains short and non-nasal; see Figure 13.5.

The greatest difference between the formant structures of nasalized and non-nasalized vowels is found in the frequency ranges of the second formants (Johnson 2003; Carignan et al., 2011). Figure 13.6 shows the F1 and F2 values of [õː], occurring before continuants, and [o], occurring before /n/ followed by a noncontinuant.
Figure 13.2 /n/ realization before continuants: transitioning CL, nagyon szalad [nɒjɒːsɒlɒd] “runs fast.”

Figure 13.3 /n/ realization before continuants: non-transitioning CL, istenség [ɪʃɛːɡ] “deity.”
Figure 13.4  /n/ realization before continuants: no CL, latinság [lɔtinɒːɡ] “the Latin people.”

Figure 13.5  /n/ realization before continuants: /n/ drop, olyan zöld [ojɒnzd] “so green.”
The distance between the first two formants in the same vowel is different, depending on whether the vowel is nasalized or not.

The realizations of /n/ before continuants as opposed to noncontinuants are markedly different. The distribution of the /n/ realizations before continuants is as follows: 49.23% with transitioning CL (Figure 13.2), 19.53% with non-transitioning CL (Figure 13.3), 14.84% without CL (Figure 13.4), and 16.4% with /n/ drop (Figure 13.5). In contrast, if /n/ is followed by a noncontinuant, the nasal consonant is retained in 98.44% of all data (1.56% are transitions).

Analysis was carried out to answer the question of whether the duration of the nasal consonant was also dependent on the following consonant. One-way ANOVA proved to be significant as well ($F(1,255)=76.535, p<0.001$). This means that the continuant or noncontinuant nature of the following consonant has an impact on the realization of /n/ (see Figure 13.7). In cases where CL is possible (before continuants), the mean duration of the /n/ realization (independent of the actual type or acoustic structure) is 25.39 milliseconds (std. dev.: 22.08 milliseconds), including all variations, while in cases where CL is not possible (before noncontinuants) the mean duration is 66.05 milliseconds (std. dev.: 22.54 milliseconds). The mean duration of the transition type of /n/ realization is 28.11 milliseconds (std. dev.: 7.54 milliseconds). Correlation analysis also supports the strong interrelation between the duration of the /n/ realization and the type of the following consonant (Pearson’s rho =0.675, $p<0.001$ at 99% confidence level), and it confirms a strong interrelation between the acoustic structure of the /n/ realization and the following consonant (Pearson's rho =0.705, $p<0.001$ at 99% confidence level). The data reveal that the nasal is shorter in all measurable cases before both continuant obstruents and approximants than before noncontinuant obstruents; see Table 13.2.

The difference in nasal duration before continuant vs. noncontinuant obstruents is significant (one-way ANOVA: $F(2,271)=48.73; p=0.001$). Post hoc Tukey tests confirm that there is no significant difference in nasal duration before continuant obstruents vs. approximants. In more careful speech styles /n/ is retained in VnC
sequences regardless of the manner of articulation of the following C. A parallel investigation of these cases by Gósy et al. (2010) found that C has a significant effect on the duration of the nasal consonant.

4.2 Vowel lengthening

All of our subjects evidence longer vowel durations before /n/ plus continuant sequences, where CL is possible, than before /n/ plus noncontinuant sequences, where CL is not possible (see Table 13.3), paralleling the findings of Gósy et al. (2010). Our goal in this section is to show the analysis of the vowel durations and provide a plausible explanation for the occurrence of vowel lengthening when /n/ is absent phonetically.

The duration differences of the vowels proved to be significant with all subjects, depending on whether the vowels occurred in contexts where CL is possible vs. where CL is not possible (one-way ANOVA: $F(255) = 18.885$, $p < 0.001$). The mean duration of vowels in the cases when the nasal is followed by approximants is 95.4 ms.

### Figure 13.7 VnC sequences: Medians and ranges of the duration of n.

### Table 13.2 VnC sequences: the duration of n.

<table>
<thead>
<tr>
<th>C types</th>
<th>Duration of n (ms)</th>
<th>Mean</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noncontinuant obstruent</td>
<td>97.8</td>
<td>30.06</td>
<td></td>
</tr>
<tr>
<td>Continuant obstruent</td>
<td>59.6</td>
<td>23.9</td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>53.25</td>
<td>20.56</td>
<td></td>
</tr>
</tbody>
</table>


The data show that vowels are about 10 milliseconds longer in CL contexts than in non-CL contexts (cf. Table 13.3). Figure 13.8 shows the medians and ranges of vowel duration values in the two contexts under consideration.

Figure 13.9 compares the duration of vowels in contexts preceding /ns/ and /nz/, revealing that the vowel durations are almost identical if /n/ is followed by continuants (casual speech contexts where /n/ generally is not pronounced). Figure 13.10 shows the differences between the duration of the vowels preceding /nz/ (/z/ representing the class of continuants) and /nd/ (/d/ representing the class of noncontinuants), revealing that there is no correlation between the two. This means that the measured vowel durations do in fact differ from each other, leading to the generalized conclusion that continuants and noncontinuants have different effects on the duration of preceding vowels. Regression analysis is a statistical tool (forecasting model) for the investigation of relationships among variables. In this analysis the coefficient of determination ($R^2$) is the proportion of variability in a data set that is accounted for by a statistical model. Regression analysis shows that the variables predict greater changes in vowel duration in contexts where CL occurs (before

<table>
<thead>
<tr>
<th>Context</th>
<th>Vowel duration (ms)</th>
<th>Mean</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL possible</td>
<td>97.62</td>
<td>20.75</td>
<td></td>
</tr>
<tr>
<td>CL impossible</td>
<td>86.65</td>
<td>19.63</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13.8  Vowel duration values in possible (DUR[+CL]) vs. impossible (DUR[-CL]) CL contexts.
Continuants) than in contexts where CL does not occur (before noncontinuants); see Figures 13.9 and 13.10.

Analyzing the data along gender class lines, there is significant difference between males and females when CL is not possible (one-way ANOVA: \( F(1,127) = 8.142, p < 0.005 \)); however, no significant difference was found when CL is possible. As seen in Table 13.4, females’ mean vowel durations are consistently longer than males.6 More precisely, the actual duration values show greater differences in the articulation of males than females in contexts where CL is not possible; cf. Figure 13.11.

The vowel plus nasal sequence is longer before continuant obstruents and approximants than before noncontinuant obstruents; see Table 13.5. This difference is statistically significant (one-way ANOVA: \( F(2,383) = 49.1; p = 0.001 \)).

Our data reveal that in VnC sequences: (a) V is shorter and /n/ is longer before noncontinuants; (b) V is longer and /n/ is shorter before continuants; and (c) the entire V+n sequence is longer before noncontinuants than before continuant obstruents7 and approximants.

\[ R^2 = 0.4727 \]

\[ R^2 = 0.0213 \]
To sum up, the actual realizations of the dento-alveolar nasal phoneme heavily depend on the following consonant, which also has an effect on the vowel preceding the nasal. The nasal consonant is stable before noncontinuants, variable before continuants.

Table 13.4  Duration values of vowels preceding /n/ depending on gender and context.

<table>
<thead>
<tr>
<th>Vowel duration (ms)</th>
<th>Females</th>
<th>Males</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
<td>Mean</td>
</tr>
<tr>
<td>[ns]</td>
<td>94.5</td>
<td>16.16</td>
<td>88.81</td>
</tr>
<tr>
<td>[nz]</td>
<td>100.64</td>
<td>15.17</td>
<td>99.03</td>
</tr>
<tr>
<td>[nʃ]</td>
<td>108.63</td>
<td>17.86</td>
<td>99.59</td>
</tr>
<tr>
<td>[nʒ]</td>
<td>93.63</td>
<td>12.83</td>
<td>96.15</td>
</tr>
<tr>
<td>[nt]</td>
<td>92.581</td>
<td>13.94</td>
<td>84.11</td>
</tr>
<tr>
<td>[nd]</td>
<td>92.86</td>
<td>8.072</td>
<td>87.36</td>
</tr>
<tr>
<td>[nds]</td>
<td>88.96</td>
<td>10.74</td>
<td>76.68</td>
</tr>
<tr>
<td>[ntʃ]</td>
<td>91.47</td>
<td>13.34</td>
<td>79.17</td>
</tr>
</tbody>
</table>

Figure 13.11  Vowel duration values of females and males in possible vs. impossible CL contexts (F = female; M = male; DUR = duration).

Table 13.5  VnC sequences: the duration of V + n.

<table>
<thead>
<tr>
<th>C types</th>
<th>Duration of V + n (ms)</th>
<th>Mean</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noncontinuant</td>
<td></td>
<td>184.45</td>
<td>28.9</td>
</tr>
<tr>
<td>Obstruent</td>
<td></td>
<td>157.22</td>
<td>35.1</td>
</tr>
<tr>
<td>Continuant obstruent</td>
<td></td>
<td>148.65</td>
<td>34.1</td>
</tr>
<tr>
<td>Approximant</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We are now in a position to explain the second component of CL: vowel lengthening. A crucial fact to bear in mind is that the lengthening of the vowel before an unpronounced /n/ is 100% correlated with nasalization: under both transitioning and non-transitioning types of CL, the vowel that is the reflex of /n/ is both long and nasalized, albeit to a lesser degree in the transitioning case (see Figures 13.2 and 13.3), underscoring the gradient nature of the process. As we have argued above, there are good articulatory reasons for /n/ not to be realized in the shape of a nasal consonant before continuant consonants. An extreme, one might say immediate, resolution of this pressure is to drop the nasal consonant lock stock and barrel, without any residue; in our experimental study, this happened in roughly 16% of the time. But more commonly, the dissolution of /n/ is not as sudden. With the nasal cavity opened up by a lowered velum, the tongue fails to make a closure in the oral cavity, bringing about an articulatory time span which is characterized as having nasal airflow but lacking oral gesture. This unarticulated nasal phase can not exist in a vacuum; it most naturally settles on and becomes part of the preceding vowel. It is this extra nasal span that gives the host vowel its newfound nasality and expanded space. The nasal tonality can be an addendum to the vowel (transitioning CL), or it can be absorbed by the vowel to various degrees (non-transitioning CL). Either way, the spatial dimension of the vowel expands.

5 Discussion

We have presented acoustical evidence, buttressed by statistical data, for a CL phenomenon in Hungarian, according to which a vowel is lengthened and nasalized (among other realizations) before /n/ plus continuant consonant sequences. We have also shown that CL is absent if the nasal consonant is followed by a non-continuant consonant. We have provided a phonetic explanation for these facts, based on the (co)articulation of the segments involved.

One might counter that the appropriate domain for the treatment and explanation of our data is phonology, rather than phonetics. Indeed, almost universally, the treatment of CL in the recent literature has been couched in phonological terms, spanning several theoretical models, of which two stand out: rule based (e.g., Wetzels and Sezer 1986; Hayes 1989; Vago 1998; Topintzi 2006, among others), principally the autosegmentally based Mora Theory, and constraint-based (e.g., Shaw 2007; Kiparsky 2011, among others), as in the various models of Optimality Theory (Prince and Smolensky 2004, and what follows). Below we highlight the basic claims of these two general frameworks as they relate to CL.

According to the basic claims of classic Mora Theory, as developed most fully by Hayes (1989), /n/ in the coda position of a syllable is associated with a mora (μ), the structural unit of phonological weight. When /n/ drops out, the principle of “mora preservation” forces the mora unit of /n/ to reassociate to the preceding short vowel, which thereby becomes bimoraic: its inherent mora is coupled with the reassociated mora from the vanished /n/. Vowels which are associated with two morae (whether inherent or derived) are interpreted as long. Schematically: \( V^\mu n^\mu \rightarrow \overline{V}^\mu \rightarrow V^\mu V^\mu = \text{long } V \).
In the constraint-based approach of classic Optimality Theory and its successor versions (Prince and Smolensky 2004; McCarthy 2002, 2007, among others), choosing the output candidate [VːC] as optimal from the input /VnC/ entails minimally the following: (1) the highly ranked markedness constraint *n CONT (/n/ cannot be followed by a continuant consonant) can be respected in a number of ways, including, but not limited to: (a) deleting /n/; (b) inserting a vowel between /n/ and the continuant consonant; (c) deleting the continuant consonant. Of the competing output candidates, option (a) is flagged as optimal by proper constraint ranking; (2) /n/ is moraic, and both its mora unit and [nasal] feature must be preserved in the most harmonic output (mora and [nasal] faithfulness); (3) a floating (unassociated) mora is associated with the preceding vowel, which becomes long by virtue of its bimoraic status.

Based on impressionistic observations, a number of characteristics of the Hungarian CL process in VnC sequences have been mentioned in the literature (Siptár and Törkenczy 2000): optionality, high degree of variability, confinement to casual speech, phrase level application, lack of structure preservation. On the whole, these characteristics find reasonable accounts in phonologically oriented models, such as Mora Theory and Optimality Theory. However, our experimental study has unearthed two additional characteristics: extensive articulatory and acoustic interplay, and most importantly, robust gradient effects. The last characteristic in particular poses a serious challenge to treating the facts of the Hungarian CL process in VnC sequences strictly in categorical terms within the confines of the phonological module. Rather, the broad range of outcomes provides textbook clues that we should look to noncategorical, phonetically based theories that are rife with insightful explanations in terms of articulation, perception, and production. Articulatory Phonology (Browman and Goldstein 1986, and what follows) is one such research program, among other phonetically based frameworks.

6 Conclusion

In this work, we have offered a phonetically based reinterpretation of a case of CL that is customarily analyzed in phonological terms. Our results are in keeping with recent research that has succeeded in providing phonetically based explanations for sound patterns that have long been relegated to the domain of phonology (see Gordon 2002; 2006; Kavitskaya 2002; Hayes et al. 2004, among others).

Acknowledgments

R. Vago’s research was supported in part by a PSC-CUNY Research Award. We are grateful to Doug Whalen for discussion.

Notes

1 For alternative interpretations of the phonetic values of the Hungarian vowels, see Vago (1980, 2006), Kenesei et al. (1998), Siptár and Törkenczy (2000), and Gósy (2004), among others.
The nasal quality that results from CL is noticeably stronger than that of vowels with normal co-articulation effects when adjacent to a nasal consonant. The nasal quality of vowels in the latter context is insignificant and will not be indicated in our phonetic transcriptions.

CL is blocked if /n/ is in syllable initial position (e.g., *latin-unk* ‘our Latin’ [lɔt.in.nuŋk] or is followed by pause (e.g., *latin* ‘Latin’ [lɛnti.ni]).

The place of assimilation of /n/ is optional across word boundary (Siptár and Törkenczy 2000). Thus, the first two entries in (2) may also be pronounced as [nɛɟɛṃpɛrʦ] and [nɒŋnəbɛnto], respectively.

The loss of nasals before fricatives (mostly voiceless, less commonly voiced) also has a perceptual basis; see Ohala and Busà (1995).

The only exception is before [ʒ]. This case is negligible, however: (a) the difference is less than 3 ms; (b) of all short vowel phonemes in Hungarian, /ɔː/ occurs least frequently in spontaneous speech (Gósy 2004: 89).

Further, the duration of both V and /n/ is affected by the voicing of the following C: both are shorter before voiceless obstruents than before voiced obstruents. See also Whalen and Beddor (1989) and Beddor (2007).

See the body of this contribution for the phonetic grounding of the *n CONT constraint.

Phrase level and structure changing applications are compatible with the postlexical level of phonology, as in the frameworks of Lexical Phonology (Kiparsky 1982, and what follows) and Stratal Optimality Theory (see for instance the treatment of CL in Kiparsky 2011).

The structure changing nature of the Hungarian CL process is shown by the following. In the lexical phonology module, short low vowels must be long if morpheme final and a suffix follows (Vago 1978, and what follow): e.g., *medve* “bear” [mɛdve], *medvē-k* “bear-pl”[mɛdve:k]; *kutyā* “dog” [kucn], *kutyā-k* “dog-pl” [kucə:k]. Note that the lengthened low vowels undergo change in quality: short low unrounded [ɛ] becomes long mid unrounded [eː] and short low rounded [ɒ] becomes long low unrounded [ɑː]. These changes are driven by structure preservation: there are no contrastive long low unrounded [ɛː] and long low rounded [ɒː] vowels.

In sharp contrast, when a low vowel lengthens as a result of CL, it keeps all of its feature values, creating noncontrastive vowel qualities: [ɛː] lengthens to [ɛː] (cf. *ötven* “50” [ɔtvɛn], *ötven-szer* “fifty times” [ɔtvɛ:sɛɾ]) and [n] lengthens to [nː] (cf. *hatvan* “sixty” [hɔtvɔn], *hatvan-szor* “sixty times” [hɔtvɔ:sɔɾ]).

These facts are compatible with a postlexical phonological analysis of CL (see Siptár and Törkenczy 2000), as well as postphonological, as advocated here.

We are cognizant of the fact that most cases of CL are high-level processes that find their motivation in diachrony or systematic phonology.

**References**


Päri Consonant Mutation as Defective Root Node Affixation

Jochen Trommer

1 Introduction

The most common approach to morphologically triggered consonant mutation in generative grammar is to derive it from the affixation of floating segmental features (Akinlabi 1996; Zoll 1996; Wolf 2005, 2007; Grijzenhout 2011; Trommer 2012). For example, the antipassive in Päri (Andersen 1988a), a Western-Nilotic language spoken by roughly 10,000 speakers in Southern Sudan, is expressed by hardening stem-final liquids (e.g., $gɛɛr \Rightarrow gɛed-o$ “build,” $kwal \Rightarrow kwʌd-o$ “steal,” Andersen 1988a:91) might be modeled by representing the antipassive morpheme as a suffix consisting of the feature [‐sonorant] which attaches to the consonant of the stem-final consonant and replaces its [+sonorant] specification in the phonological output. However, from a more global perspective, the antipassive is atypical for consonant mutation in Päri and in Western Nilotic more generally, which typically involve – at least for some input consonants undergoing mutation – a bisegmental output. Thus of the four general types (“grades”) of consonant mutation discussed by Andersen (1988a) only grade 2° (which includes the antipassive) leads to strictly monosegmental outputs. All other types of consonant mutation produce either nasal + stop clusters or double consonants. In Anywa (Reh 1993), virtually all cases of mutation apart from the antipassive also lead to the doubling (gemination) of underlying consonants. Even cases of mutation which do not exhibit overt bisegmental outputs often behave as if they consisted of more than one consonant. For example, the Mayak antipassive (Andersen 1999) blocks an otherwise general process of intervocalic stop lenition, a fact which can be naturally captured if the antipassive is a defective consonant which shields the stem consonant from an intervocalic environment at the relevant level of representation. This suggests that
consonant mutation in Western Nilotic is not due to underlying floating features, but to the affixation of full segments (consonants), which phonologically merge with stem-final consonants along the lines suggested for consonant mutation more generally in Bye and Svenonius (2012) and de Lacy (2012).

In this chapter, I will develop an analysis of consonant mutation in the verbal morphology of Päri under the assumption that all cases are due to affixation of (partially defective, i.e., featurally underspecified) segments. Päri presents probably the most fascinating array of consonant mutation patterns in Western Nilotic since it employs most of the patterns (stopping, lenition, gemination and formation of nasal+stop clusters) which are only partially attested in other languages. The remaining parts of this section introduce the theoretical framework I will assume (subsection 1.2), and the relevant Päri data (subsection 1.1). Section 2 presents the detailed analysis, and section 3 develops the arguments that the defective segment approach to Päri is superior to analytic alternatives.

1.1 Consonant mutation in Päri

Päri is a Western-Nilotic language of the Northern Lwoo sub-branch, spoken by roughly 10,000 speakers in Southern Sudan. It has a rich non-concatenative morphology crowded on monosyllabic stems employing changes in tone, vowel quality, segmental features of consonants, and length. Morphemes are generally monosyllabic, where lexical roots end in single consonants and show contrasts in vowel length (or have diphthongs), whereas affix syllables are open and allow only single short vowels. Moreover, there are no complex onsets or codas apart from combinations of single onset consonants with the following glide [w]. Segmental affixation is restricted to maximally a single suffix and a single prefix per word. All data in this chapter are from the fieldwork of Andersen (1988abc, 1989), especially Andersen (1988a). Although Andersen’s descriptions pair descriptive thoroughness with analytic depth, our empirical understanding of Päri phonology and morphosyntax is clearly fragmentary. In particular, Andersen does not systematically describe nominal morphology, the consequences of morphological derivations for tone and vowel quality, and many details of segmental phonology. Thus, where necessary, I will conjecture on morphophonological properties of Päri from the raw paradigms Andersen provides and from Päri’s closest genetic relative, Anywa, which is documented in more detail in Reh (1993). (1) gives typical examples of consonant mutation for transitive verbs.

All categories of consonant mutation are concomitant with vowel suffixes (the prefix a- in the Ø, 3SG, 3PL, and frequentative forms is the exponent of completive aspect).

(1) Päri consonant mutation – examples (Andersen 1988a: 89–92, 97, 98)

<table>
<thead>
<tr>
<th>1.0°</th>
<th>2.0°</th>
<th>2.1°</th>
<th>3.0°</th>
<th>4.1°</th>
<th>5°</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-kat</td>
<td>a-kad-e</td>
<td>kad-o</td>
<td>a-kat-e</td>
<td>kann-o</td>
<td>a-kánd-i</td>
</tr>
<tr>
<td>a-jap</td>
<td>a-jab-e</td>
<td>jabol-o</td>
<td>a-jap-e</td>
<td>jamm-o</td>
<td>a-jaamb-i</td>
</tr>
<tr>
<td>a-kwaan</td>
<td>a-kwaan-e</td>
<td>kwaan-o</td>
<td>a-kwaand-e</td>
<td>kwaan-o</td>
<td>–</td>
</tr>
</tbody>
</table>
Consonant Mutation as Defective Root Node Affixation

I use the following abbreviations: AP = antipassive, BEN = benefactive, FC = focus, FQ = frequentative, INC = inchoative. “Ø” stands for non-derived forms without inflection. To enhance readability, I omit the tone diacritics provided by Andersen.

(1) does not include grade 4.0° examples because only intransitive verbs exhibit this grade. Verbs with underlying long vowels do not employ grade 5°, but 3.0° in the frequentative; see section 3.2 for discussion. There is one more mutation type not included in (1) (grade 3.1°) which I will not discuss in this chapter because it simply combines antipassive morphology/mutation transparently with grade 3.0° mutation. All other gaps in (1) are accidental gaps in the data set Andersen provides for specific verbs. (2) provides a full list of the basic “grades” (patterns) of consonant mutation for all phonemic consonants of Päri (note that the only obstruents in Päri are stops).

(2a) shows the raw data in the notation of Andersen (1988a) adapted to IPA symbols, and (2b) the phonological representations I assume as the output of Stem Level Phonology. Here and in the following, I adopt Andersen’s claim that phonetically long consonants are phonologically sequences of two segments. Note also that double consonants in Päri never occur in underived words. They are always the result of consonant mutation. (2b) abstracts away from intervocalic Word-Level lenition ([j] ⇒ [jj], [g] ⇒ Ø, [ɟɟ] ⇒ [jj]). This process can be observed in a similar way in the better described phonological system of Anywa (Reh 1993). I also take the mutated [r] in grade 3.0° as an instance of [ɟɟ] at the Stem Level which is lenited to [jj] at the Word Level. Again the same process can be observed in Anywa. Since Päri neutralizes length contrasts in root-final obstruents at the Word Level (“Voiceless stops are short after long or diphthongal stem vowels, and long after short stem vowels,” Andersen 1988a: 71), Andersen’s notation of oral stops in grade 3.0° is analytically arbitrary, and I notate these also as double consonants because this allows for a more consistent characterization of 3PL-exponence as consistently doubling underlying consonants. Grade 4.1° results from applying an allomorph of the benefactive affix to the regular output of antipassive formation (grade 2.0°), hence there are no relevant input liquids – stem-final liquids are mutated into stops by antipassive mutation (for grades 2.0°, 3.0°, and 4.1°, there are additional verbal categories which employ these grades):

(2) Grades of consonant mutation in Päri

a. Andersen’s (1988a) raw data

\[
\begin{array}{cccccccc}
\text{Grade} & 1° & 2.0° & 2.1° & 3.0° & 4.0° & 4.1° & 5° \\
p & b & b & p & mm & mm & mb \\
\dd & d & d & \dd & nn & nn & nd \\
\end{array}
\]
1.2 The framework: Stratal Colored Containment Theory

My analysis is based on a version of Optimality Theory which is close to the original implementation of the theory proposed by Prince and Smolensky (1993) in adopting hierarchical autosegmental representations and the Containment Assumption on candidate generation, but which adopts the representation of epenthesis by morphological colors from Colored Containment Theory (van Oostendorp 2006; Revithiadou 2007), and generalizes the Containment Assumption to association lines (Radical Containment). Following, McCarthy and Prince (1993), Myers (1997), Kiparsky (2000), Bermúdez-Otero (2003, 2012), I assume a stratal organization of OT, generalized to single morpheme evaluation as in Trommer (2011).
**Morphological colors and epenthesis.** Following van Oostendorp (2006) and Revithiadou (2007), I assume that morphological structure is minimally reflected in phonological representations by coloring. At the interface of morphology and phonology, every morpheme M of an underlying representation UR is assigned a unique color C (i.e., a color which is distinct from all other colors C⁰ in UR), and every phonological component (i.e., every node and every association line) of M is also assigned C. (3) illustrates coloring with two hypothetical morphemes a and l. Color is notated here and throughout this chapter by background boxes with distinctive shading. (3b) is a candidate based on the input (3a) which adds epenthetic [i], syllables, a mora, and epenthetic association lines. (3c) shows the same candidate in a slightly different notation which highlights the epenthetic character of association lines by denoting them with broken lines.

Colors have two important consequences for phonological computation. First, they allow us to distinguish underlying material from the epenthetic one: epenthetic material is colorless (in (3): black), and the Containment Assumption GEN does not permit changing or removing the color of underlying material. Second, they make it possible to determine whether two phonological elements belong to the same morpheme or not. Crucially colors do permit phonological constraints to distinguish morphemes, but not to identify them. Thus coloring phonology cannot determine whether l in (3) is a 3SG affix or a noun root. It just “knows” that it is a morpheme which is distinct from a. Coloring thus grants the bare minimum of accessibility phonology may have to morphological structure.

**Containment and possible operations of GEN.** I endorse the Radical Containment assumption that phonological material can never be literally removed from phonological input representations in the course of phonological computation (Prince and Smolensky 1993; van Oostendorp 2008: 1365). The candidate-generating function GEN is thus restricted to the following changes it may perform on underlying forms (phonetic visibility is conceived as an elementary attribute of association lines, but not of phonological nodes):

(4) Possible operations of GEN.

a. Insert epenthetic nodes (prosodic nodes, feature nodes, segmental root nodes) or phonetically visible association lines between nodes
b. Mark an underlying association line as phonetically invisible.
(4a) implements the implicit assumptions held on containment and GEN in the earliest version of Optimality Theory (Prince and Smolensky 1993), whereas (4b) replaces deletion of association lines by a less invasive operation: marking for phonetic invisibility (indicated in the following by “=”). (5b) shows some representative candidates generated by GEN for the input in (5a). (5b ii, iii) contain epenthetic association lines licensed by (4a). In (5b, iii), the association line between the second \( \mu \) and [e] is marked as phonetically invisible according to (4b).

(5) Candidates generated by GEN

a. Input:

b. Candidates: (i) \( \sigma \)

   (ii) \( \mu \)

   (iii) \( \mu \)

   [e]

   [ie]

   [i]

Deletion as phonetic non-realization. (Non)pronunciation of underlying material is implemented as phonetic noninterpretation of phonological material following the axioms in (6). These axioms in (6) are equivalent to an obligatory version of the operation Stray Erasure in pre-OT and early OT-phonology (Steriade 1982; Itô 1988; Prince and Smolensky 1993).

(6) Axioms of phonetic realization

a. Nodes: A phonological node is phonetically realized if and only if it is dominated by the highest prosodic node of the candidate through an uninterrupted path of phonetically visible association lines.

b. Association lines: An association line is phonetically realized if and only if it is marked as phonetically visible and connects two phonetically realized nodes.

“Highest” in (6a) refers to the familiar prosodic hierarchy… Prosodic Word > Foot > \( \sigma \) > \( \mu \) > (Nespor and Vogel 1986) under the assumption that GEN does not generate candidates which contain more than one highest prosodic node (say two \( \sigma \)-nodes, but no Foot or PWord nodes). Thus the highest prosodic node in all examples of (5b) is the \( \sigma \)-node because the representations do not contain foot or word nodes. (7b) shows the part of (5b, iii) repeated as (7a) which is spelled out by phonetic interpretation. [e] and the second \( \mu \) of (7a) are not in (7b) because the upper association line through which they are dominated by \( \sigma \) is phonetically invisible. (5b, i) a slightly different way of deletion: [i] and the first \( \mu \) are not phonetically interpreted because they are not dominated by \( \sigma \) at all. In the following, I will often
emphasize non-pronunciation by gray shading, thus the i in (7b) would appear as i (see the table in (14) for examples).

(7)

In the following, I will call phonetically realized substructures such as (7b) the “P(honetic) Structure” or simply “P” of a candidate, the substructure of a candidate which corresponds to the input “M(orphological) Structure” or “M,” and the overall candidate its “I(ntegrated) Structure” or “I” (7a).

Markedness constraints and the Cloning Hypothesis. Following Trommer (2011), I propose that markedness constraints are subject to the Cloning Hypothesis formulated in (8):

(8) The Cloning Hypothesis

Every markedness constraint has two incarnations, a phonetic (P-structure) clone and a general (I-structure) clone:

- The general clone refers to complete phonological representations. The phonetic clone refers to the phonetically visible substructure of phonological representations.

The relation of clones proposed here is parallel to the relation of different types of faithfulness constraints in Correspondence Theory (McCarthy and Prince 1994, 1995), for example MaxIO (for the input-output relation), and MaxBR (for the base-reduplicant relation). Whereas these constraints are structurally identical, they refer to different subrepresentations of candidates (or more exactly of input candidate mappings), but can be ranked independently in individual grammars. This is also true for the markedness clones developed here. There are no M-Structure clones of constraints. I illustrate the Cloning Hypothesis with a constraint which plays a crucial role in the analysis of Päri vowel length polarity discussed in section 3.2, the ban on trimoraic vowels. The phonetic clone of the constraint in (9b) is a constraint which is uncontroversially unviolated in most languages of the world, and blocks phonetically extra-long vowels (or in other words a three-way vowel length contrast). The general clone in (9a) generalizes this constraint to the full phonological structure. The phonetic clone is marked here, as throughout the chapter by underlining, whereas the general clone does not have any explicit marking.

(9) a. *V^{3\mu} → Assign * to every V which is dominated by more than 2 $\mu$s (in I).

b. *V^{3\mu} → Assign * to every V which is dominated phonetically by more than 2 $\mu$s (in P).
To see where (9a) and (9b) substantially differ, consider the case of a floating mora affix which is attached to a base with a bimoraic vowel (10a). The structure in (10b) which results from straightforward association of the floating mora to appropriate base nodes violates both, \( ^*V^3\mu \) and \( ^*V^3\mu \), since all nodes and association lines here are phonetically interpreted (they are dominated by the highest prosodic node, \( \sigma \)). On the other hand, (10c) does not violate \( ^*V^3\mu \) because in the subrepresentation of the structure which is phonetically interpreted, (10c), the vowel is only associated to two \( \mu \)s. On the other hand, (10c) still violates \( ^*V^3\mu \) because in its overall structure the vowel is associated to three moras. Thus the general version \( ^*V^3\mu \) effectively blocks association of the affix-\( \mu \) to a bimoraic vowel even under deassociation of other moras, a point which is crucial for the derivation of vowel-length polarity in Päri (see section 3.2 for discussion).

\[
\begin{array}{c}
\text{(10)} \\
\begin{array}{c}
\mu & \mu \\
\sigma & \mu \\
\end{array} \\
\begin{array}{c}
\text{a.} \\
\mu \\
\end{array} & \\
\begin{array}{c}
\mu & \mu \\
\sigma & \mu \\
\end{array} \\
\begin{array}{c}
\text{b.} \\
\mu \\
\end{array} & \\
\begin{array}{c}
\mu & \mu \\
\sigma & \mu \\
\end{array} \\
\begin{array}{c}
\text{c.} \\
\mu \\
\end{array} & \\
\begin{array}{c}
\mu & \mu \\
\sigma & \mu \\
\end{array} \\
\begin{array}{c}
\text{c'.} \\
\mu \\
\end{array}
\end{array}
\]

Phonological strata. In line with recent developments in the phonological literature, I assume that OT is organized in part derivationally (Kiparsky 2000, McCarthy 2008, 2010, Inkelas and Caballero 2013). In particular, I will assume the stratal organization of OT-grammar proposed in Trommer (2011), where all lexical roots are evaluated in isolation at Morpheme-Level optimization, stems (roots with all eventual Stem-Level affixes of a word form) at the Stem Level, and complete words (including all Word-Level affixes if there are any) at the Word Level. Thus every lexical root passes through three subsequent optimization cycles reflecting increasing morphological complexity. What sets the stratal architecture apart from alternative derivational OT-models is that the number of levels is universally fixed, and that different levels may assume different constraint rankings in the same language. Crucially, since mutation inducing morphemes in Western Nilotic are Stem-Level affixes, their effects are often opaque – obscured by later Word-Level processes.

Again following Trommer (2011), I assume that affixes, just as roots, also undergo independent OT-evaluation before they are concatenated (Stem-Level affixes at the Morpheme Level, and Word-Level affixes at the Stem Level), accounting for effects, where word-sized (e.g., bisyllabic) affixes pattern with roots (Bermúdez-Otero 2013), but also capturing morpheme-structure constraints, for example the fact that roots are typically required to be fully prosodified (to form a prosodic word), whereas affixes in many languages are restricted to syllabic or subsyllabic size. The morphological properties of a morpheme (such as the root-affix distinction) are only visible at the level where this morpheme is first/independently evaluated, which leads to a specific version of the Indirect-Reference Hypothesis: in morphologically complex forms, the effects of morphological features is indirect (e.g., due to different
prosodification of roots and affixes). Thus Trommer (2011) reduces the root-dominant vowel Stem-Level harmony system of Päri to prosodic differences derived at the Morpheme Level: lexical roots are required to show full prosodification at Morpheme-Level evaluation, and conversely affixes are required to remain under foot size (see Downing 2006 and references there for evidence that morpheme status is linked in this way to prosodic size). At the subsequent Stem-Level evaluation, harmony is purely sensitive to prosody, maintaining the [ATR] values of stressed vowels in roots and assimilating unstressed affix vowels to the quality of the main stressed vowel, without directly accessing the morphological difference between the formatives.

2 Analysis

The stratal architecture allows capturing grades 1° and 2.0° as the result of general phonological voicing alternations at the Stem Level; hence they do not exhibit morphological consonant mutation at all. The focus marker -a is a Word-Level affix, therefore focus forms pattern at the Stem Level with bare stems (Ø), while all other grades apart from 1° exhibit at the Stem-Level vowels immediately following the root-final consonants. Given these observations, the voicing of non-word initial obstruents in Päri is completely predictable by their phonological environment at the Stem Level: obstruents in word-final position (grade 1°) and in stop clusters (cf. grade 3.0°) are consistently voiceless, while intervocalic single stops (grades 2.0° and 2.1°) and stops after nasals are voiced (cf. grades 3.0° and 5°). Thus voicing is only contrastive in word-initial position. I assume that these generalizations are due to the constraint ranking Ident $^{P\text{wd}}_{\pmvc}$ (“Retain the underlying voicing of PWord-initial segments”) $>$ (ROR)$_{\pmvc}$ (“An obstruent must phonetically share the [+vc] feature of surrounding sonorants”) $>$ $^*$D (“No phonetic voiced obstruents”) at the Stem Level.

Also mutation for manner features is severely constrained by exceptionless phonological restrictions. Thus, as noted by Andersen, the coda-onset combinations in (10b) provide an exhaustive list of the available options for intervocalic consonant clusters in the language more generally. In particular, the following generalizations hold: (i) Päri has neither complex onsets nor codas (hence the maximum number on non-nuclear elements between two syllable peaks is 2). The only exception to this statement is the glide [w] which may occur in stems after another onset consonant as in kwa:n, “count” (Andersen 1988a: 89). I assume that this option is restricted to the position preceding the head of a PWord (the stem vowel), hence it is not an option for the affixal/post-stem material under discussion here. (ii) All coda-onset clusters share their place and [±continuant] features, and (iii) the only option for two segments in this context not to share all manner features is a coda nasal followed by a (homorganic) stop. In the following, I will assume that this strict phonotactic template is enforced by a set of undominated phonological constraints which I will summarily abbreviate as $\sigma$-Cond. $\sigma$-Cond is closely related to the Coda Condition of Itô
(1988), which captures the observation that in many languages marked place features in codas are restricted to geminates and nasals which share their place specification with a following (onset-initial) stop.

I assume that all Päri affixes inducing consonant mutation start either (1) with a single consonantal root node unspecified for place, (2) with two such root nodes, or (3) with a root node specified for place. I will start my discussion with the first group which comprises the majority of verbal affixes in Päri.

2.1 Affixation of a single placeless •-node

(11) lists representative morpheme entries for affixes which start with a single segmental root node that is underspecified for place in the Distributed-Morphology notation for vocabulary entries (Halle and Marantz 1993). Since standard DM does not countenance portmanteau affixes, the antipassive:benefactive exponent is interpreted here as the spellout of a benefactive head in the context of an antipassive head. I omit additional affix consonants because they are irrelevant for mutation. “•” is the symbol I will use in the following to represent root nodes. Following Padgett (1995), Wolf (2005, 2007) and against McCarthy (1988), I assume that root nodes do not inherently specify class features.

(11) Single placeless •-affixes.

\[
\begin{align*}
\text{a.} & \quad \text{3PL} \leftrightarrow \begin{array}{c}
-\text{son} \quad 3.0^\circ \\
\cdot
\end{array} \\
\text{b.} & \quad \text{INC} \leftrightarrow \begin{array}{c}
+\text{son} \quad 4.0^\circ \\
\cdot
\end{array} \\
\text{c.} & \quad \text{BEN/AP} \leftrightarrow \begin{array}{c}
\text{nas} \quad 4.1^\circ \\
\cdot
\end{array}
\end{align*}
\]

The core of the analysis consists now of the basic faithfulness constraint \text{Max} \text{•} (“Root nodes which are in M should also be in P”) and the association constraint \text{•} \rightarrow \text{pl(ace)} (“Every root node should be associated to a place node in I”). \text{•} \rightarrow \text{pl} is the generalized version of the standard autosegmental requirement that every segment has a specification for place (this is equivalent to \text{HAVEPLACE} in Itô and Mester 1993; Lombardi 1999; Walker 2000; Parker 2001; Beckman 2004; McCarthy 2008). In tandem, these constraints have the effect that the underspecified floating root node is fully integrated into the phonological overall structure by sharing (dominating) the \text{pl}-node of the preceding consonant, and being dominated by the affixal \text{σ}-node. Coda consonants in Päri (12a) are apparently non-moraic, therefore I assume that they are dominated by the \text{μ} of the preceding vowel (cf. Broselow \textit{et al.} 1997 and Bermúdez-Otero 2001):
Now, whereas both segments appear in the output, they cannot do so unchanged; for many consonant combinations, manner features of one segment extend to the other one, to satisfy Pāri phonotactics encoded by σ-COND. An obvious and exceptionless generalization on the resulting clusters is that a [nasal] feature sponsored by one of the input consonants also shows up in the output consonant cluster (to verify this, check the BEN:AP column and all rows for input nasals in (2b)). This is encoded by the faithfulness constraint (13a) ([n] abbreviates in the following [nasal], [±s] [±sonorant], and [±c] [±consonantal]). For all consonant combinations without an underlying nasal, but a stem-final glide, the output is the double version of this glide, which follows straightforwardly from ranking Max below Max\textsubscript{\text{\text{c}}}. For all other consonant combinations, the manner features of the affix consonant are realized faithfully. I assume that this is a positional faithfulness effect (Beckman 1997). The faithfulness constraint in (13c) specifically protects the [±sonorant] value of onset consonants.
For some combinations, the clash of two consonants also leads to the insertion of additional featural material not sponsored by any of the input segments. All relevant cases (affixation of inchoative: [+son] to stems ending in stops, and of 3PL: [‐son] to [l] final stems) involve insertion of the feature [nasal] in underlying consonants and concomitant insertion of manner features (supporting the claim by Rice 1993 and related work that nasals are the default realization of sonorants). I assume that this is due to undominated Dep constraints for other sonorant manner features, Dep [lat(eral)] Dep [rhot(ic)], and Dep [‐cons].

It should be easy to see that this results in the correct outputs for all stem nasals. For the BEN:AP affix with an underlying [nasal] consonant (11c) we get maximally faithfully two output nasals. The same output results for the [+son] consonant of the inchoative affix (11b) because homorganic nasals are the only sonorants allowed in Päri after other nasals, and the [‐son] of the 3PL (11a) transparently results in a nasal + stop cluster. All these outputs leave the stem nasal unaltered and extend the affix consonant without changing any of its underlying specifications.

Underlying stem-final stops require more drastic adjustment. Whereas the affixation of [‐son] (of the 3PL affix) to a stop results transparently in a double stop, Päri phonotactics do not allow for combinations of a stop and a following sonorant. Here the constraints in (13) get decisive, as shown in (14) and (15), where ([F1][F2])F3 abbreviates two consonantal •-nodes, underlyingly associated to the features F1 and F2 respectively and sharing the features F3. The affix nasal/sonorant remains unchanged due to Maxn (14) and MAX•↓s (15), while the stem obstruent takes over the manner features of the affix sonorant. The sonorant in grade 4.0° in (15) is also realized as a nasal since creating other sonorants would violate undominated Dep [lat] or Dep [rhot] (example forms in tables are taken, where not otherwise noted, from the table in (1) and abstract away from all morphophonological alternations not directly related to mutation morphology; (15) is based on mt (1.0°) ⇒ minno (4.0°) “be delicious,” Andersen 1988a: 94). Recall that shading indicates non-pronunciation.

(13) Max constraints on manner features in Päri and their ranking

a. Maxn Assign * to every [nasal] node in M which is not in P

b. Maxc Assign * to every [‐cons] node in M which is not in P

c. *Max↑ Assign * to every onset •-node in P which dominates a [±son] node in M but does not dominate it in P

d. Max↑ Assign * to every •-node in P which dominates a [±son] node in M but does not dominate it in P

For some combinations, the clash of two consonants also leads to the insertion of additional featural material not sponsored by any of the input segments. All relevant cases (affixation of inchoative: [+son] to stems ending in stops, and of 3PL: [‐son] to [l] final stems) involve insertion of the feature [nasal] in underlying consonants and concomitant insertion of manner features (supporting the claim by Rice 1993 and related work that nasals are the default realization of sonorants). I assume that this is due to undominated Dep constraints for other sonorant manner features, Dep [lat(eral)] Dep [rhot(ic)], and Dep [‐cons].

It should be easy to see that this results in the correct outputs for all stem nasals. For the BEN:AP affix with an underlying [nasal] consonant (11c) we get maximally faithfully two output nasals. The same output results for the [+son] consonant of the inchoative affix (11b) because homorganic nasals are the only sonorants allowed in Päri after other nasals, and the [‐son] of the 3PL (11a) transparently results in a nasal + stop cluster. All these outputs leave the stem nasal unaltered and extend the affix consonant without changing any of its underlying specifications.

Underlying stem-final stops require more drastic adjustment. Whereas the affixation of [‐son] (of the 3PL affix) to a stop results transparently in a double stop, Päri phonotactics do not allow for combinations of a stop and a following sonorant. Here the constraints in (13) get decisive, as shown in (14) and (15), where ([F1][F2])F3 abbreviates two consonantal •-nodes, underlyingly associated to the features F1 and F2 respectively and sharing the features F3. The affix nasal/sonorant remains unchanged due to Maxn (14) and MAX•↓s (15), while the stem obstruent takes over the manner features of the affix sonorant. The sonorant in grade 4.0° in (15) is also realized as a nasal since creating other sonorants would violate undominated Dep [lat] or Dep [rhot] (example forms in tables are taken, where not otherwise noted, from the table in (1) and abstract away from all morphophonological alternations not directly related to mutation morphology; (15) is based on mt (1.0°) ⇒ minno (4.0°) “be delicious,” Andersen 1988a: 94). Recall that shading indicates non-pronunciation.

(14) Stop + nasal (4.1°) – kan.no ‘plaint CP:AP’

<table>
<thead>
<tr>
<th>Input:</th>
<th>[‐s][+s]</th>
<th>(kat.No)</th>
<th>σ-Cond</th>
<th>Maxn</th>
<th>Maxc</th>
<th>*Max↑</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>([‐s][+s])pl.m</td>
<td>(kan.no)</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>([‐s][+s])pl.m</td>
<td>(kat.to)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>([‐s][+s])pl</td>
<td>(kat.no)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(13) Max constraints on manner features in Päri and their ranking

a. Maxn Assign * to every [nasal] node in M which is not in P

b. Maxc Assign * to every [‐cons] node in M which is not in P

c. *Max↑ Assign * to every onset •-node in P which dominates a [±son] node in M but does not dominate it in P

d. Max↑ Assign * to every •-node in P which dominates a [±son] node in M but does not dominate it in P
Päri Consonant Mutation as Defective Root Node Affixation

(15) Stop + sonorant (4.0°) – minno (4.1°) ‘be delicious INC’

<table>
<thead>
<tr>
<th>Input: [-s][+s]</th>
<th>σ-Cond</th>
<th>Max_n</th>
<th>Max_c</th>
<th>oMax_*</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ([-s][+s])_{PL,M}</td>
<td>(mit.no)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ([-s][+s])_{PL,M}</td>
<td>(mit.to)</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>c. ([-s][+s])_{PL}</td>
<td>(mit.no)</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

Stems with underlying final glides (i.e., [-consonantal] segments) retain these due to high ranked Max_c for [-son] and [+son] affixes (16), but give way to the [nasal] specification of the BEN:AP affix under the force of undominated Max_n (17) (I mark epenthetic material here in bold; [±t] abbreviates [±continuant]).

(16) Approximant + obstruent (3.0°) – laaww-ε ‘wash 3PL’

<table>
<thead>
<tr>
<th>Input: [+s–c][−s]</th>
<th>σ-Cond</th>
<th>Max_n</th>
<th>Max_c</th>
<th>oMax_*</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ([+s–c n][−s])_{PL,M}</td>
<td>(laam.be)</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ([+s–c][−s])_{PL,M}</td>
<td>(laap.pe)</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>c. ([+s–c][−s])_{PL,M}</td>
<td>(laaw.we)</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. ([+s–c][−s])_{PL}</td>
<td>(laab.be)</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(17) Approximant + nasal (4.1°) – dooŋŋ-ο ‘weed CP:AP’

<table>
<thead>
<tr>
<th>Input: [+s–c+t][n]</th>
<th>σ-Cond</th>
<th>Max_n</th>
<th>Max_c</th>
<th>oMax_*</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ([+s–c+t n][−t])_{PL,M}</td>
<td>(dooŋŋ.o)</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ([+s–c+t][n−t])_{PL,M}</td>
<td>(dooj.jo)</td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. ([+s–c+t][n−t])_{PL}</td>
<td>(dooj.ŋo)</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stem-final liquids lead to straightforward doubling with [+son]:INC ([l] ⇒ [ll], [r] ⇒ [rr]), which requires neither deletion nor insertion of manner features, but leads to complications with [-son]:3PL because [l].stop and [r].stop are not

(18) [l] + obstruent (3.0°) – kwand-ε ‘steal 3PL’

<table>
<thead>
<tr>
<th>Input: [+s+1][−s]</th>
<th>σ-Cond</th>
<th>Max_n</th>
<th>Max_c</th>
<th>oMax_*</th>
<th>Max_*</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ([+s+1 n][−s])_{PL}</td>
<td>(kwand.ε)</td>
<td></td>
<td></td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>b. ([+s+1][−s])_{PL,M}</td>
<td>(kwat.ε)</td>
<td></td>
<td></td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c. ([+s+1][−s])_{PL,M}</td>
<td>(kwal.lε)</td>
<td></td>
<td></td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>d. ([+s+1][−s])_{PL}</td>
<td>(kwal.de)</td>
<td></td>
<td></td>
<td>*!</td>
<td>*</td>
</tr>
</tbody>
</table>
permitted by $\sigma$-COND (18d). Turning both consonants into sonorants is blocked by onset prominence ($^0$Max•↓s), and a double stop (18b) by its general counterpart Max•↓s (which is too low ranked to be relevant for the consonant combinations discussed so far).

Without further provisos, we expect the same output for stem final [r] which however results in a double stop ([ɟɟ]). I propose that the output [ɲɟ] (see Trommer 2011 for arguments that [r] is phonologically palatal, not alveolar) is blocked in this case by the constraint $^*_r$pln formulated in (19):

(19) $^*_r$pln: Assign $^*$ to every PL which is associated to [+rhot] and [nas] segments in I.

$^*_r$pln is a highly general containment-based formulation of the ban on nasal rhotics. Whereas the P-version of this constraint would just block segments which are phonetically nasal and rhotic, the I-version in (19) also penalizes changes of nasals into rhotics (or vice versa). Thus a process such as $r \Rightarrow n$ would induce a violation because the coronal feature is linked to a root node which is associated (morphologically to [+rhotic]) and (accidentally the same) root node which is linked (phonetically) to [nasal]. $^*_r$pln is defined with respect to a pl-node instead to a segmental root so that it also penalizes cases where two segments in a feature sharing relationship form a kind of partial complex segment such that one of them is linked (morphologically or phonetically) to [nasal], and the other one (morphologically or phonetically) to [+rhotic]. In (20) it is shown that undominated $^*_r$pln blocks the transformation of the rhotic into a nasal. The only way to maintain the [–son] of the affix -•, and to satisfy $\sigma$-COND is to turn [r] into an oral stop:

(20) [r] + obstruent (3.0°) – geeʃʃɛ ‘build 3PL’

<table>
<thead>
<tr>
<th>Input: [+s+r][–s] (geer.Te)</th>
<th>$^*_r$pln</th>
<th>$\sigma$-COND</th>
<th>Max•</th>
<th>Max•c</th>
<th>$^*$Max•↓s</th>
<th>Max•↓s</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ([+s+r n][–s]$_{pl}$) (geep. je)</td>
<td>$^*$!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ([+s+l][–s]$_{pl,m}$) (geej. je)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ([+s+r][–s]$_{pl,m}$) (geer. re)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>d. ([+s+r][–s]$_{pl}$) (geer. je)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2 Affixation of 2 pl-less •-nodes in the frequentative

All mutation patterns discussed so far render stem-final consonants uniformly either more stop like (grade 3.0°) or more sonorous/nasal (grades 4.0°/4.1°). The frequentative (grade 5°) combines both effects and results for almost all input consonants in a homorganic nasal-stop sequence. In the analysis developed here, this follows naturally if the frequentative affix is represented by two pl-less consonantal root nodes, the first one [+son] and the second one [-son]:

-
Affixation of (21) to a verb leads to an inherent structural problem: since all Pāri verb roots have an underlying final consonant or glide. We get three consonantal root nodes, but σ-COND allows only two nonsyllabic intervocalic elements. The net result of the constraint ranking developed so far is that one of the consonantal root nodes must be deleted (22a). That it is the stem-final consonant which yields follows from the standard condition on hierarchical phonological representations (with respect to pl-nodes) that non-root nodes should not be dominated by more than one root node, where “root node” is to be understood graph theoretically as a node that is not dominated by any other node (since the stem-μ node of the stem and the σ-node of the affix in (22a) are in turn part of a single prosodic word, this is the only root node of the overall structure and the only root node dominating pl, thus satisfying *rpln):

(22) pl-Ursupation in the Frequentative (grade 5°)
This is a case of what Zimmermann (2013) calls phonological “usurpation”: phonologically defective affixes usurp stem material and trigger thus indirectly deletion of segments in the stem. Usurpation follows from crucially undominated * → PL and *\(^\circ\)PL\(^\circ\). The only way to satisfy both constraints for the defective affix *-nodes is to fully integrate them phonetically, that is, to pronounce them, whereas rendering the association lines of the stem-final C phonetically invisible (resulting directly in its non-phonification) does not violate these constraints since they are already morphologically associated, and containment preserves this association in I. Thus the Containment Assumption and the generalized versions of autosegmental constraints lead to a strategic structural advantage for defective (underlyingly only partially integrated) phonological material in the competition for phonetic realization under phonotactic pressure.

For most input consonants the surviving affix *-nodes maintain their underlying manner specifications resulting in the standard spellout of [+son] by a nasal (due to DEP [lat] and DEP [rhot]) and of the [-son] node as a stop, but in a few cases the phonotactic constraints and faithfulness constraints introduced above interfere. For stem final [r], undominated *\(_r\)PL\(_n\) blocks associating its PL-node to a cluster containing a nasal (23c), but all other two * clusters containing a sonorant are also excluded, laterals by DEP [lat], [r.r] (23b) because it violates oMax \(^\circ\), and by \(\sigma\)-Cond \([r]\) cannot be combined with any other segment in a cluster. Thus the only remaining option is stopping of the entire cluster (23a) (Andersen does not give a concrete example for this pattern):

(23) \((\text{[r]}) + \text{Sonorant} + \text{obstruent (Grade 5\(^\circ\)})\)

<table>
<thead>
<tr>
<th>Input: ([+s+r][+s] [–s]) (r.R.T)</th>
<th>*PL(_n)</th>
<th>(\sigma)-Cond</th>
<th>Max(_n)</th>
<th>Max(_c)</th>
<th>oMax (^\circ)</th>
<th>Max (^\circ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (([+s+r][+s] [–s]))(_{pl,m}) (j.j)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. (([+s+r][+s] [–s]))(_{pl.m}) (r.r)</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. (([+s+r][+s] [–s]))(_{pl}) (n.j)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

Whereas Andersen does not provide data or an explicit statement for stem-final glides, the prediction the constraint ranking makes is that the output is the same as for grade 3.0\(^\circ\) mutation. Max\(_c\) enforces survival of [–cons] resulting in double glides:

(24) \(\text{Approximant} + \text{sonorant} + \text{obstruent (Grade 5\(^\circ\)})}\)

<table>
<thead>
<tr>
<th>Input: ([+s–c][+s] [–s]) (r.R.T)</th>
<th>*PL(_n)</th>
<th>(\sigma)-Cond</th>
<th>Max(_n)</th>
<th>Max(_c)</th>
<th>oMax (^\circ)</th>
<th>Max (^\circ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (([+s–c][+s] [–s]))(_{pl,m}) (m.m)</td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. (([+s–c][+s] [–s]))(_{pl.m}) (w.w)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>c. (([+s–c][+s] [–s]))(_{pl}) (m.b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

There is a fascinating empirical complication to the frequentative mutation pattern: Verbs with underlyingly long vowels or diphthongs do not show the consistent nasal+ stop cluster, but double consonants for stem-final input stops. Thus for verb
roots with two vocalic μs, consonant mutation in the frequentative is identical to the grade 3.0° pattern we have observed for the 3PL (see section 3.2 for discussion of the effects frequentative morphology has on vowel length and diphthongs):

(25) Grade alternation in the frequentative (Andersen 1988a:89)

<table>
<thead>
<tr>
<th>a. Short input stems: Grade 5°</th>
<th>b. Long input stems: Grade 3.0°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>FQ</td>
</tr>
<tr>
<td>(i) a-jap ‘open’</td>
<td>a-jaamb-1</td>
</tr>
<tr>
<td>(ii) a-ŋaŋ ‘suck’</td>
<td>a-ŋoŋg-1</td>
</tr>
<tr>
<td>(iii) a-jik ‘make’</td>
<td>a-jıŋg-1</td>
</tr>
<tr>
<td>(iv) a-kat ‘plait’</td>
<td>a-kająnd-1</td>
</tr>
<tr>
<td>Ø</td>
<td>FQ</td>
</tr>
<tr>
<td>(i) a-łuop ‘speak’</td>
<td>a-łup-1</td>
</tr>
<tr>
<td>(ii) a-łukk ‘wash’</td>
<td>a-łök-1</td>
</tr>
<tr>
<td>(iii) a-rit ‘sew’</td>
<td>a-rit-1</td>
</tr>
<tr>
<td>(iv) a-poot ‘beat’</td>
<td>a-poot-1</td>
</tr>
</tbody>
</table>

I assume that the different behavior of bimoraic and monomoraic verbs at the Stem Level is a combined effect of prosodic word minimality at the Morpheme Level and positional faithfulness at the Stem Level. At the Morpheme Level, bimoraic roots form a syllable under the pressure of Parse Segment (Ps-Seg), a foot, and a concomitant prosodic word (26). On the other hand, for monomoraic roots, high ranked Foot-Binarity (Ft-Bin) blocks formation of feet and PWds, as shown in (27) (parentheses indicate syllables, and square brackets PWds; recall that all Päri morphemes are monosyllabic and that coda consonants are not moraic).

(26) Morpheme Level: Evaluation of a bimoraic root

<table>
<thead>
<tr>
<th>Input: lʊμμp</th>
<th>Ps-Seg</th>
<th>Ft-Bin</th>
<th>Ps-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>lʊμμp</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td></td>
<td>![lʊμμp]</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>![lʊμμp]</td>
<td>![lʊμμp]</td>
<td>![lʊμμp]</td>
<td>*!!</td>
</tr>
</tbody>
</table>

(27) Morpheme Level: Evaluation of a monomoraic root

<table>
<thead>
<tr>
<th>Input: jaμp</th>
<th>Ps-Seg</th>
<th>Ft-Bin</th>
<th>Ps-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>![jaμp]</td>
<td></td>
<td>![jaμp]</td>
<td>*!!</td>
</tr>
<tr>
<td>![jaμp]</td>
<td>![jaμp]</td>
<td>![jaμp]</td>
<td>*</td>
</tr>
</tbody>
</table>

At the Stem Level, the high ranked positional faithfulness constraint Max↓ requires that a PWd-final segment which dominates [-sonorant] in M also does so in P, excluding PL-usurpation as in (28c) (see Krämer 2003 for discussion of positional faithfulness effect at the right word edge). This results in a double stop (28a) (association with the second affix consonant) since association of stem final PL with the first affix consonant, which is [+son], violates σ-Cond (28b):
(28) Emergence of root-final • with bimoraic verbs (grade 5°)

\[
\text{Input: } (= 28-e) \quad \sigma \text{-COND} \quad ^c\text{MAX} \quad \text{MAX} \quad \text{MAX} \quad \downarrow \quad \text{PL}
\]

\[
\begin{array}{|c|c|c|c|c|}
\hline
& \mu & \mu & \sigma & V & V
\hline
\text{a.} & V & V & \cdot \cdot & \cdot \cdot & V
\hline
\text{b.} & V & V & \cdot \cdot & \cdot \cdot & V
\hline
\text{c.} & V & V & \cdot \cdot & \cdot \cdot & V
\hline
\text{d.} & V & V & \cdot \cdot & \cdot \cdot & V
\hline
\text{e.} & V & V & \cdot \cdot & \cdot \cdot & V
\hline
\end{array}
\]

\[\text{MAX} \] is crucially ranked below \(^c\text{MAX} \), hence in grades 4.0° and 4.1° of bimoraic verb roots it is still the manner features of the affix (providing the onset consonant) that prevail. Thus the effect of \( \text{MAX} \) is effectively confined to the frequentative forms of bimoraic stop-final verbs. In turn, \(^c\text{MAX} \) ensures that the configuration in (28b) cannot be saved by rendering the first affix-• [-son].

2.3 Affixation of a pl-specified •-node

Antipassive (grade 2.1°) mutation seems now to lead to a paradox: if all consonant mutation is affixation of (partially defective) segmental root nodes, and high ranked constraints guarantee the realization of exactly two intervocalic consonants, how can we capture a consonant mutation process that involves not doubling, but minimal feature modification of a single segment? The solution I propose lies again in the containment-based generalization of markedness constraints, especially the constraint \( ^c\text{C}_{n} \), which is the generalized version of the phonological constraint encoding the cross linguistic markedness of complex segments:
Assign * to every consonantal root node which is associated to more than 1 PL-node in I.

Now assume that the exceptional representational property of the antipassive affix is that its initial consonant – in contrast to all other Stem-Level affixes of Päri is preassociated to a PL-node, CORONAL as shown in (30):

Affix consisting of a •-node specified for PL

The representation in (30), \( \text{pl}^*C_{\text{pl}} \), and the constrained phonotactic options of Päri conspire to restrict the antipassive to monoconsonantal outputs, as shown in (31). The straightforward realization of both intervocalic consonants (31e) is excluded because \( \sigma\-\text{COND} \) only allows consonant clusters which share a PL-feature, but PL-sharing of two segments (31d) violates \( \text{pl}^*C_{\text{pl}} \) just in the same way as a complex segment which

Liquid stopping in the Antipassive (grade 2.1°)

<table>
<thead>
<tr>
<th>Input: = (31-f)</th>
<th>( \text{pl}^*C_{\text{pl}} )</th>
<th>( \sigma-\text{COND} )</th>
<th>( \text{IO}^*\text{PL} )</th>
<th>( \text{MAX} ) •</th>
<th>( \downarrow \text{PL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ( \mu )</td>
<td>( \text{PL} )</td>
<td>( \text{COR} )</td>
<td>(m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ( \mu )</td>
<td>( \text{PL} )</td>
<td>( \text{COR} )</td>
<td>(t)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ( \mu )</td>
<td>( \text{PL} )</td>
<td>( \text{COR} )</td>
<td>(tp)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. ( \mu )</td>
<td>( \text{PL} )</td>
<td>( \text{COR} )</td>
<td>(m.p)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. ( \mu )</td>
<td>( \text{PL} )</td>
<td>( \text{COR} )</td>
<td>(m.t)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. ( \mu )</td>
<td>( \text{PL} )</td>
<td>( \text{COR} )</td>
<td>(m)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
phonetically realizes the pl-features of both input consonants (31c) (note that “sharing” here means technically to be associated to the same pl-feature token not just association to the same type, hence a sequence such as [n.t] would be impossible if [n] and [t] are linked to different Cor nodes). Thus the only viable option is to delete one of the underlying consonants (31a, b). Note that root node deletion in (31a, b) violates $\bullet \rightarrow \text{pl}$ (which hence must be dominated by $\text{pl}^\bullet \text{C}_\text{pl}$), but not $^\circ \text{pl}^\circ$ (the only root node dominating the affix-pl in (31a) is the affix-$\bullet$, the only root node dominating the other pl-nodes in (31a, b) is the PWord of the overall structure).

Obviously, the affix consonant shows up with stem-final liquids (which are in turn deleted as in (31b)), whereas all other final stem consonants surface faithfully at the cost of the affix consonant (31a). This follows in essential respects from the constraint ranking already established. Stem-final nasals are retained by undominated $\text{Max}_n$, and stem-final glides due to $\text{Max}_c$. On the other hand, the preference for realizing segmental (non-)sonorancy of the affix consonant that we have observed in the bisegmental mutation patterns above is obviated in the antipassive since the affix consonant in (31a) is not in onset position, hence does not violate $^\circ \text{Max}^\circ$. I assume that the decision between (31a) and (31b) for all other types of stem-final consonants is due to the low-ranked faithfulness constraints $\text{Max}_s$, $\text{Max}_\text{DORSAL}$, $\text{Max}_\text{LABIAL}$, which effectively protect stem-final stops (32), and otherwise the coronal obstruent of the antipassive affix (33). Since these constraints are ranked below all other faithfulness constraints on features/$^\circ$- nodes discussed so far, they do not have any effect in the bisegmental mutation cases. ($^\circ \text{Max}^\circ$) is omitted in (32) and (33): its only potential effect is to protect base final stops of bimoraic verb roots, thus simply reinforcing the effect illustrated in (33) that all base final stops are retained in the antipassive).

(32) Deletion of affix [t] after stem [k] (2.1° antipassive)

<table>
<thead>
<tr>
<th>Input: k.t</th>
<th>$\text{Max}_n$</th>
<th>$\text{Max}_c$</th>
<th>$^\circ \text{Max}^\circ$</th>
<th>$\text{Max}_s$</th>
<th>$\text{Max}_\text{DOR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. k</td>
<td></td>
<td></td>
<td>$^\circ$</td>
<td>*</td>
<td>*!</td>
</tr>
<tr>
<td>b. t</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*!</td>
</tr>
</tbody>
</table>

(33) Deletion of stem [r] before affix [t] (2.1° antipassive)

<table>
<thead>
<tr>
<th>Input: r.t</th>
<th>$\text{Max}_n$</th>
<th>$\text{Max}_c$</th>
<th>$^\circ \text{Max}^\circ$</th>
<th>$\text{Max}_s$</th>
<th>$\text{Max}_\text{DOR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. r</td>
<td></td>
<td></td>
<td>$^\circ$</td>
<td>*</td>
<td>*!</td>
</tr>
<tr>
<td>b. t</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*!</td>
</tr>
</tbody>
</table>

3 Discussion/alternatives

As pointed out by Bermúdez-Otero (2012), the major problem that non-concatenative morphology posits for phonological theory is not that there are too few possible analyses, but that there are too many. Consider for example Päri 3PL mutation,
which I have analyzed here as affixation of an underspecified consonantal root node that is part of the underlying affix specification. This pattern could also be analyzed as affixation of the floating feature [−sonorant] in the 3PL-affix which associates to stem-final stops without phonetic effect, but leads to insertion of an additional consonantal root node after sonorants (see Zoll 1996 for a similar analysis of floating laryngeal affixation in Yowlumnee). Assuming that codas in Päri are moraic, a third possibility is that the changes in the 3PL are actually triggered by affixing a floating μ that is associated to a coda nasal, where possible, and otherwise deleted, where insertion of stops would serve the sole purpose of “shielding” the nasal coda from being resyllabified as the onset of the following affix-σ. Against this space of conceivable analyses, it is clearly desirable to prune the analytic options learners of natural languages have if they face consonant mutation. Here I want to defend the position that all consonant mutation is the result of affixing segmental root nodes. Positive evidence for this claim results if it can be shown that neither floating feature affixation nor μ-affixation alone can account for consonant mutation in a given language.

In the following, I will argue that none of these alternatives is sufficient to capture the full array of consonant mutation patterns in Päri.

3.1 Consonant mutation by affixation of floating features

Affixation of floating features is the most frequent approach to consonant mutation in the literature and an obvious analytic option for patterns as the Päri antipassive where underlying mono segmental consonants remain single segments after undergoing mutation. An important result of the analysis developed here is the constructive proof that even this type of pattern receives a natural implementation under the assumption that all consonant mutation is driven by affixes of segments. Conversely, floating feature affixation provides no principled explanation why consonants in other morphological contexts of Päri lengthen/double. Consider for example the 3PL. Assuming that consonant mutation in this case is triggered by affixation of the floating feature [−son] could capture the mutation of nasals (and laterals) into nasal + stop sequences. The bisegmental realization could be interpreted naturally as the result of the conflict to maintain the underlying featural structure of the nasal intact and to realize [−son] without inserting an additional pl-feature. However this type of approach does not explain why glides in the 3PL undergo doubling. Outputs such as [jj] and [ww] do not realize [−son] in any obvious way. Similarly, the sonorizing effect of grade 4.0° inchoative on stem-final stops could be captured by affixing floating [+son], but this would not account for the consistent consonant doubling triggered in this context. A single nasal would realize [+son] just as well as the double one; both [n] and [n,n] involve the overwriting of the underlying [−son] of the stem consonant; moreover single nasals are phonotactically perfectly possible in intervocalic position and in fact show up in other consonant grades and underived verbs. Thus the floating feature approach does not provide a natural approach to doubling in most of the consonant mutation patterns of Päri. Finally it is unclear how a pure floating feature approach would capture the fact that one hardening process in a given
language generates monosegmental output consonants (the Päri antipassive) and other one bisegmental outputs (e.g., the 3PL forms). This follows in the defective •-affixation analysis from the simple fact that consonant affixation results in deletion of a consonant in specific contexts. •-affixation also accounts naturally for a further fact that remains mysterious under a floating feature analysis. Hardening of [r] leads to a palatal stop in the 3PL [j], but to a coronal stop ([d]) in the antipassive. If this were due to a floating feature (say [-anterior]) we would incorrectly expect that the same contrast shows up at least in the BEN:AP forms for stem final [t]. On the other hand, in the root node affixation analysis, [r] and [j] are consistently palatal whereas the [d] in the antipassive is simply the faithful realization of the affix consonant.

3.2 Consonant mutation by μ-affixation

Insertion of nasals and gemination are among the classical examples for which μ-affixation analyses have been proposed in the literature (see Samek-Lodovici 1992; Davis and Ueda 2006). Thus, as shown in (34), both creation of nasal + stop clusters (34a) and gemination (34b) in Päri might be attributed to the affixation of floating moras in grades 3.0°, 4.0°, and 5° since they provide obvious ways to integrate a μ into the coda of a preceding σ, while guaranteeing an onset to the following syllable (which fails with simple onsets or codas as in (34c, d)).

However, μ-affixation by itself cannot explain the segmental changes which are concomitant to gemination (the nasalization of stops in grades 4.0°/4.1° and the stopping or nasalization of liquids in grade 3.0°). Moreover, without additional assumptions, μ-affixation does not explain why the floating μ results in some morphological contexts in gemination (34b), and in other ones in a nasal + stop complex (34a), even where the underlying stem consonants are identical (compare e.g., nasal final verbs in grades 3.0° and 4.0°). Augmenting μ-affixation by further floating features, one might assume that grade 3.0° is underlyingly an affix consisting of a floating μ, and the feature [-sonorant], grade 4.0° a floating μ + the floating feature [+sonorant], and grade 2.1° (the only grade without gemination) a bare floating [sonorant], in parallel to the affix specifications under the defective root node analysis stated in (11) and (30), but this would inherit a basic problem of the pure feature affixation approach discussed in section 3.1: we would expect that [-sonorant] in grade 2.0° mutates [r] into the palatal stop [j], in parallel to the [r] ⇒ [ hà ] patterns in grades 3.0° and 5.0°. To endow the antipassive/grade 2.1° affix by a further floating place feature ([‐anterior]) that obligatorily overwrites the putative [+anterior] specification of [r] would again not solve the problem, since then we would expect that this also overwrites underlying stem [t] that would result in [c] instead of the correct output [t]. This is not to say that the μ-affixation analysis could not be made to work by auxiliary assumptions such as further underlying material or by stipulating different morphologically triggered constraints (indexed constraints in the sense of Pater 2007, 2009, or constraint rankings, cf. Inkelas and Zoll 2005; Jurgec 2010 on cophonologies). The crucial point is that output differences in the output shapes is expected anyway if consonant
mutation is due to different underlying consonants, but requires additional assumptions under μ-affixation. A third problem for μ-affixation is that it lets us expect that single morphological categories lead to homogeneous changes to the underlying weight of base verbs, but this expectation is thwarted by the fact that weight of morphologically derived verbs in Päri is independently subject to orthogonal non-concatenative processes affecting vowel length. Compare for example benefactive and 3PL morphology, which both exhibit grade 3.0° consonant mutation.

The 3PL forms can in fact be interpreted as integrating an additional μ as in (34a, b), but the benefactive also shows consistent shortening of vowels or monophthongization of diphthongs (the benefactive also shifts [-ATR] stem vowels to [+ATR], a further non-concatenative process triggered by many morphological categories in Päri that is

---

**Table 34** Consonant mutation as μ-affixation

<table>
<thead>
<tr>
<th>Input: (34-e)</th>
<th>Max μ</th>
<th>ONS</th>
<th>NoCoda</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>a.</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>b.</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>c.</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>d.</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
<tr>
<td>e.</td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
<td><img src="#" alt="Diagram" /></td>
</tr>
</tbody>
</table>
in principle independent of consonant mutation and shortening). In fact, under the standard assumption that vowels are inherently moraic and long vowels bimoraic, a natural analysis of morphological vowel shortening is affixation of a floating \( \mu \), which for phonological reasons cannot be fully integrated into the phonological structure of its base. Thus under the analysis proposed for Anywa in Trommer (2014), shortening is triggered by a moraic suffix which associates to the \( \sigma \)-node of the base, but not to any segmental root node (consonants cannot bear a \( \mu \), and association to the base vowel would lead to crossing association lines). If the base already has two \( \mu \)s, one of these is deassociated to maintain the restriction to maximally bimoraic syllables (cf. the constraint \( \forall \mu \mu \) in section 1.2).

(36) Shortening by \( \mu \)-suffiation (\( \! \mu = \text{AP} \))

<table>
<thead>
<tr>
<th>Input:</th>
<th>Output:</th>
<th>Blocked:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma )</td>
<td>( \sigma )</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>( \mu ) ( ! \mu )</td>
<td>( \mu ) ( ! \mu )</td>
<td>( \mu ) ( ! \mu )</td>
</tr>
<tr>
<td>V C</td>
<td>V C</td>
<td>V C</td>
</tr>
<tr>
<td>a. 1( \mu )-Base</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>( \sigma )</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>( \mu ) ( \mu ) ( ! \mu )</td>
<td>( \mu ) ( \mu ) ( ! \mu )</td>
<td>( \mu ) ( \mu ) ( ! \mu )</td>
</tr>
<tr>
<td>V C</td>
<td>V C</td>
<td>V C</td>
</tr>
<tr>
<td>b. 2( \mu )-Base</td>
<td>( \Rightarrow )</td>
<td>( \Rightarrow )</td>
</tr>
</tbody>
</table>

This approach to vowel shortening is perfectly compatible with the analysis of consonant mutation in Päri developed above. Coda consonants are taken to be
no n-moraic in Päri, and do hence neither influence vowel length nor are they influenced by vowel shortening. Hence the benefactive affix that shows 3.0° mutation + shortening comprises a floating μ, and a [-son] consonant, whereas the 3PL (which exhibits 3.0° mutation, namely consonant lengthening, but not vowel shortening) has an underlying [-son] consonant, and lacks floating moraic material.

At a more descriptive level, benefactive morphology is problematic for a purely μ-based analysis because it does not lead to a uniform weight effect of its base. Whereas a short vowel root exhibits one more consonantal μ and keeps vowel length constant, long vowel roots compensate the lengthening in coda position by shortening the vowel, and keep their overall μ-count constant. On the other hand under the analysis sketched in (36), where codas are interpreted as consistently non-moraic, the uniform effect of benefactive affixation is a decrease of moraic weight by one. Note also a more fundamental problem with the μ-affixation analysis in (34): it requires the assumption that in the bases to affixation, codas are non-moraic, whereas they bear moras in affixed forms.

There is a second set of data, where the morphological manipulation of vowel length is concomitant to consonant mutation – vowel-length polarity: in the frequentative, underlyingly short base vowels become long (37a), and underlying long vowels become short (37b).

(37) Vowel-length polarity in the frequentative (Andersen 1988a: 89)

<table>
<thead>
<tr>
<th></th>
<th>FQ</th>
<th></th>
<th>QQ</th>
<th>QQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Short stem vowels: lengthening</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) a-jap ‘open’</td>
<td>a-jaamb-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ii) a-ŋɔ ‘suck’</td>
<td>a-ŋɔɔŋ-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iii) a-jík ‘make’</td>
<td>a-jíiNg-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iv) a-ŋɔt ‘plait’</td>
<td>a-ŋɔɔt-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Long stem vowels: shortening</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) a-loop ‘speak’</td>
<td>a-loop-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ii) a-kwaan ‘count’</td>
<td>a-kwand-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iii) a-ŋɔ ‘sew’</td>
<td>a-ŋɔ-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iv) a-waŋ ‘burn’</td>
<td>a-waŋ-1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Independently from the vowel-length effect, the frequentative is the morphological pattern which prima facie lends itself the most to an analysis of consonant mutation by μ-affixation since the mutation pattern for stop-final stems is sensitive to the length of verb roots (cf. section 2.2). A possible argument might be that affix μs preferentially associates to stem vowels lengthening short base vowels. If nasal + stop clusters are taken as non-moraic, their emergence in this context can be interpreted as a strategy to avoid a trimoraic σ. On the other hand, the affix-μ cannot attach to long base vowels and leads instead to gemination of the base final stop, where gemination in turn leads to shortening of the base vowel (again due to the avoidance of trimoraic syllables). This line of attack works for stem-final stops, but not for sonorants because it predicts that stems with long vowels and nasals should not exhibit vowel shortening (*a- kwaand-1 should be well formed). This points to a general problem with the idea that consonant mutation and vowel-length polarity are both a consequence of μ-affixation. The vowel-length effect is completely independent of the realization of stem final consonants. This follows again from adopting the analysis of vowel-length polarity in Anywa from Trommer (2014) to Päri: polarity is
triggered by a $\mu$-prefix which associates to short base vowels (and the correspondent $\sigma$) resulting in lengthening, but only to the $\sigma$-node for long base vowels to avoid a vowel linked to three $\mu$s. Owing to an additional generalized ban on trimoraic syllables, one of the base $\mu$s is disassociated resulting in shortening. Note also crucially that the moraic frequentative exponent in contrast to the shortening benefactive-$\mu$ is a prefix which allows it to access short base vowels without line crossing.

(38) Length polarity by $\mu$-prefixation ($\mu$= FQ)

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
<th>Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. 1$\mu$-Base</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) V C</td>
<td>(ii) V C</td>
<td></td>
</tr>
<tr>
<td>b. 2$\mu$-Base</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) V C</td>
<td>(ii) V C</td>
<td></td>
</tr>
</tbody>
</table>

Again the approach which restricts the effect of $\mu$-affixation to vowels is perfectly compatible with the analysis of consonant mutation developed in section 2.2.

3.3 Consonant mutation by resizing (Pycha 2008)

Pycha (2008) argues that consonant mutation in Päri instantiates two more general strategies of non-concatenative morphology: phonetic “upsizing” and “downsizing” along scales as in (39). Thus she treats grade 2.0° as downsizing – since voicing voiceless stops renders them shorter, just as the intervocalic deletion of [k] which applies in this grade, but most other grades as upsizing. Thus grade 5° obviously makes base verbs longer by extending stem-final single consonants into bisegmental nasal + stop sequences.

(39) Down Up

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>kad</td>
<td>kat</td>
</tr>
</tbody>
</table>

The central appeal of Pycha’s approach is that it allows the capturing of the morphological equivalence of phonologically heterogeneous operations such as obstruent voicing and deletion. However, if the analysis of Päri developed in this chapter is on the right track, both downsizing and upsizing follow from simple and general grammatical properties: the co-occurrence of voicing and deletion in downsizing follows from the fact that obstruents are voiced intervocically ($t \Rightarrow d$, $k \Rightarrow g$) at the Stem Level, which feeds in turn intervocalic deletion at the Word Level, two
Päri Consonant Mutation as Defective Root Node Affixation

3.4 Päri consonant mutation and the status of segments

As far as the analysis presented here is correct, it not only provides evidence for the viability of a mainly segmental approach to non-concatenative morphology, but also to a specific conception of segments. At least under the restrictive computational framework provided by Containment Theory, the Päri data imply that segments do not inherently specify a fixed set of features, namely segments are bare root nodes. To see this point, consider the possibility to reconstruct the analysis in the feature geometric model of McCarthy (1988) which is based on the assumption that all segmental root nodes are specified for the features [±consonant] and [±sonorant]. Under this assumption, the 3PL (3.0°) affix could not be a [-sonorant] root node because 3.0° mutation leads to doubling even in cases where [-son] cannot surface as with underlying approximants (jj). Since in McCarthy’s approach [±son] is also found only in root nodes (i.e., it cannot occur on a tier different from the root node tier), [-son] can also not be understood as a floating feature. If Päri approximants are interpreted as moraless vowels (but see Staroverov 2014 and references cited there for arguments that specific approximants are [+consonantal]), the same argument extends to the feature [±consonant]. Thus the Päri mutation data provide evidence that root nodes are at least not the inherent host for [±sonorant], and possibly also not for [±consonant], contra to McCarthy’s approach. However, in a theory where segmental root nodes are not uniformly specified for any specific features, namely that they are objects without inherent substantial labels, their function in
phonological representation becomes almost indistinguishable from one occupied by the timing units of early autosegmental theory, namely the X-slots proposed by Levin (1985), or, if glides in Päri turn out to be [+cons], the Cs and Vs originally proposed by McCarthy (1979, 1981): the representation of every (singleton) sound contains exactly one root node/X-slot, and the representation of (at least some) consonantal geminates contain two. This position is advocated for a timing slot model by Ringen and Vago (2011), and for a feature geometric model with respect to root nodes in Selkirk (1990), the approach I have taken here (see Muller 2001; Topintzi 2008 for arguments that there are two types of long consonants, some associated to a single root node, but two prosodic/syllabic positions, and some associated to two-root nodes). In fact, the major difference between Vago and Ringen and Selkirk is not so much whether gemination is captured by double root nodes or skeletal timing units, but in the role attributed to moras, which are abandoned by Vago and Ringen, but retain a crucial role in Selkirk in accounting for weight and compensatory lengthening. Again the Päri data shed light on this problem: if, as argued above, floating moras are responsible for non-concatenative effects on vowel length, and defective root nodes for gemination and consonantal mutation effects in the language, this provides evidence for the richer model advocated by Selkirk which combines a two-root approach to geminates with a moraic approach to weight.

4 Summary

In this chapter, I have argued that the intricate bipositional consonant mutation pattern employed by the verbal morphology of Päri can and should be captured, not by underlying floating features, but by affixation of “defective” (underspecified) segmental root nodes. I have shown that Stratal Colored Containment Theory (Trommer 2011, 2014, 2015) provides a natural framework to account for the complex differences between different mutation triggering affixes as well as for different realizations of the same affixes for different bases (e.g., gemination of stem final stops vs. formation of a nasal + stop cluster for underlying liquids). The analysis thus provides evidence for the viability of a segmental approach to subsegmental morphophonological processes (de Lacy 2012; Bye and Svenonius 2012), and of a two-root representation of consonantal length (Selkirk 1990).

Acknowledgments

The research documented in this paper was generously funded by the DFG-grants TR 521/5-1 and TR 521/6-1. I thank Alfred Peet for crucial support, Eric Rainy and Chuck Cairns for valuable criticism, and the audience of the CUNY Conference on the Segment for helpful discussion of the material in this chapter. All remaining errors are my own.
References


15

Templates as Affixation of Segment-sized Units
The Case of Southern Sierra Miwok

Eva Zimmermann

1 Introduction

In Southern Sierra Miwok (Freeland 1951; Broadbent 1964; Sloan 1991), suffixes can require the preceding stem to conform to a certain shape. A first illustrating example is given in (1) where four different forms all based on the same verb stem “to hunt” are given. The stem is followed by different suffixes and it surfaces in a different shape in every context: It has a medial geminate in (1a), no geminate in (1b), a light open second syllable in (1c), and a long vowel in the first syllable in (1d).

(1) Templates in Southern Sierra Miwok (Sloan 1991: 152–254)
   a. hal:ik-iH-h:Y-ʔ ‘he used to hunt’
   b. halik-meh-nY-haHk- ʔe-ʔ ‘I was hunting on my way’
   c. halki-paH ‘a good hunter’
   d. ha:lik- ʔe:-nY ‘to hunt along the trail’

In her investigation of syllable structure and templates of Southern Sierra Miwok, Sloan (1991) argues that three bisyllabic templates are particularly interesting since they require an analysis assuming (partly) syllabified X-slots in the representation of morphemes. The templates in question all consist of a light syllable followed by a heavy syllable (=LH) that is either closed by a coda consonant or has a long vowel. In Sloan’s analysis, these differences are predicted from the nature of the final segmental X-slot that is either associated to the nucleus node, directly to the syllable node or is floating. In contrast, I argue for an analysis of the three LH templates that is based on the affixation of moras and underspecified segments. The analysis is therefore situated in the line of research termed “generalized nonlinear affixation”
by Bermúdez-Otero (2012) that strives to derive all instances of non-concatenative morphology without any additional assumptions simply from affixation of non-linear phonological representations that are independently motivated. The analysis also relies heavily on the insightful argumentation in Bye and Svenonius (2012) that the “independently motivated components of syntax and phonology […] do the work necessary for morphology” (p. 428).

The chapter is structured as follows: I begin with some necessary background assumptions about templates in Southern Sierra Miwok in general and the three specific LH templates that are the focus of my analysis in section 2.1. In section 2.2, crucial phonological background especially about the stress system of the language is given. In section 3, I present my optimality-theoretic analysis for the three LH templates that is crucially based on the two theoretical mechanisms of moraic overwriting (section 3.1) and realization of underspecified segments (section 3.2). Section 4 discusses the nature of these underspecified segments in greater detail. I conclude in section 5.

2 The data: LH templates in Southern Sierra Miwok

Sierra Miwok is one of five moderately diverse Miwok languages (Penutian) that can be subdivided in the three regional dialects of Northern, Southern, and Central Sierra Miwok. Southern Sierra Miwok (=SSM) was spoken over much of Mariposa Country, in the foothills of the Sierra Nevada and has only a few semispeakers or passive speakers today (Hinton 1994; Golla 2011). My data for SSM are mainly from Broadbent (1964) that is also the base for the theoretical work in Sloan (1991). Another source I rely on is Freeland (1951) (written in 1936) that focuses on Central Southern Miwok but contains informations on Northern and Southern Sierra Miwok as well. Up to now, I am aware of only one other theoretical analysis for templates in Sierra Miwok and that is on the Central variety (Bye and Svenonius 2011). Their analysis is quite similar to my own theoretical proposal based on the affixation of moras and root nodes, although the Central Sierra Miwok data they analyze is different from the three LH templates I focus on.2

2.1 The status of templates in SSM

Freeland (1951) identifies three different kinds of suffixes in SSM: first, those that do not affect the stem, including some “loosely attached” elements (= postclitics) and many derivational suffixes, second, those that trigger lengthening of the final syllable of the stem, and third, those suffixes that require a certain rhythmic pattern for the base that precedes them. In this chapter, I focus on the latter type of affixes that I call “template-requiring” in the following. Note that I use the term “template” purely descriptively to refer to fixed sequences of long/short vowels and consonants. Most verbal affixes are of the template-requiring type. For illustration, some exemplifying
template-requiring suffixes and the base form they require is given. (In the following, the numbers in parentheses refer to the page in Sloan (1991) where the respective example can be found).

(2) Examples for template-requiring affixes in SSM (Broadbent 1964; Sloan, 1991)

<table>
<thead>
<tr>
<th>Suffix</th>
<th>Gloss</th>
<th>Template requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>-h</td>
<td>‘transitional’</td>
<td>CVC</td>
</tr>
<tr>
<td>-lVmh</td>
<td>‘to be ready to…’</td>
<td>CVCCV</td>
</tr>
<tr>
<td>-iH</td>
<td>‘habitual’</td>
<td>CVC:VC</td>
</tr>
<tr>
<td>-peH</td>
<td>‘agentive’</td>
<td>CVCVC</td>
</tr>
<tr>
<td>-j</td>
<td>‘verbalizer’</td>
<td>CVCV:</td>
</tr>
</tbody>
</table>

The existence of affixes that do not require any change on their preceding base is crucial since it allows one to determine an underlying form for every stem. This distinguishes the template effects in Miwok from templatic morphology in for example, Semitic morphology (for discussion and literature see, e.g., Bat-El 2011). Such instances of “template-requiring affixes” are also attested in Yawelmani, another Penutian language of California (Archangeli 1984, 1991).

In the following, I concentrate on three classes of suffixes requiring the preceding base to conform to an LH template. These three different template-requiring suffixes are evidence that both the reference to segments and prosodic structure is necessary to account for the full range of template effects. All three of them require that the preceding base is bisyllabic and starts with a light CV syllable. Affixes of class I require that the second syllable is closed as is illustrated in (3I) with the agentive suffix -peH. Class II affixes on the other side require a long final vowel as for example the suffix ɨ in (3II). Bases preceding suffixes of class III end either in a CVC or CV syllable (3III). This last class of suffixes is hence especially interesting since it predicts one of two alternating templates that conform to the LH restriction.

(3) Examples of LH-requiring affixes

I. affix –peH “agentive” (Sloan 1991: 172)
   a. halik-peH ‘hunter’
   b. ʔokoj-peH ‘a nurse’
   c. liwaP-peH ‘speechmaker’

II. affix –ɨ “do what is characteristic of . . . ” (Sloan 1991: 177)
   d. wiliː-ɨ ‘to flash, of lightening’
   e. puluː-ɨ ‘to dip up’
   f. moliː-ɨ ‘shade’

III. affix –na “benefactive” (Sloan 1991: 173)
   g. kojow-na ‘to tell for someone’
   h. hekaː-na ‘to clean for someone’
   i. tetiː-na ‘to gather for someone’
The variation between CVC and CV in the forms preceding class III suffixes is bound to the number of underlying stem consonants. Three-consonantal stems as in (3g) surface as CV.CVC whereas stems with only two consonants in their underlying representation (3h, i) surface as CV.CV. The table in (4) shows this different behavior of two- and three-consonantal stems with some examples.\(^3\)

(4) followed by followed by followed by

<table>
<thead>
<tr>
<th>Biconsonantal stems</th>
<th>class I affix</th>
<th>class II affix</th>
<th>class III affix</th>
</tr>
</thead>
<tbody>
<tr>
<td>ko:l (147)</td>
<td>kolu?</td>
<td>kolu:</td>
<td>kolu:</td>
</tr>
<tr>
<td>ho:ja (147)</td>
<td>hoja?</td>
<td>hoja:</td>
<td>hoja:</td>
</tr>
<tr>
<td>liw:a (172 + 175)</td>
<td>liwa?</td>
<td>liwa:</td>
<td>liwa:</td>
</tr>
<tr>
<td>pel:e (171)</td>
<td>pele?</td>
<td>pele:</td>
<td>pele:</td>
</tr>
<tr>
<td>hek:a (173)</td>
<td>he ka:</td>
<td>heka:</td>
<td>heka:</td>
</tr>
<tr>
<td>mol:i (177)</td>
<td>moli:</td>
<td>moli:</td>
<td>moli:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Three-consonantal stems</th>
<th>class I affix</th>
<th>class II affix</th>
<th>class III affix</th>
</tr>
</thead>
<tbody>
<tr>
<td>polat (147)</td>
<td>polat</td>
<td>pola:</td>
<td>polat</td>
</tr>
<tr>
<td>helaj (177)</td>
<td>helaj</td>
<td>hela:</td>
<td>helaj</td>
</tr>
<tr>
<td>wikis (169)</td>
<td>wikis</td>
<td>wikis</td>
<td>wikis</td>
</tr>
<tr>
<td>halik (172)</td>
<td>halik</td>
<td>halik</td>
<td>halik</td>
</tr>
<tr>
<td>pult (177)</td>
<td>pulu:</td>
<td>pulu:</td>
<td>pulu:</td>
</tr>
<tr>
<td>wili:p (177)</td>
<td>wili:</td>
<td>wili:</td>
<td>wili:</td>
</tr>
</tbody>
</table>

Interestingly enough, the three LH templates therefore result in only two different surface structures (CV.CVC and CV.CV:) that are distributed differently for two- and three-consonantal stems.

(5) The three LH templates

<table>
<thead>
<tr>
<th>Biconsonantal stem</th>
<th>Three-consonantal stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I requires</td>
<td>CV.CVC</td>
</tr>
<tr>
<td>Class II requires</td>
<td>CV.CV:</td>
</tr>
<tr>
<td>Class III requires</td>
<td>CV.CV:</td>
</tr>
</tbody>
</table>

A close look at the examples in (3) makes it apparent that different phonological strategies apply to ensure that the stem conforms to the form requirements of the three different templates. Instances of CV-metathesis (M), realization of an additional \(\breve{i}\) (i), realization of an additional ? (?)\(^4\), V-shortening (S), C-deletion (D), V-lengthening (L), and degemination (G) can be found. The table (6) compares different stems with a respective LH form ending in a closed syllable (I/III) or a long vowel (class II/III) and the various phonological changes that apply.
Phonological changes

<table>
<thead>
<tr>
<th>stem</th>
<th>CVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>halki</td>
<td>halik ✓</td>
</tr>
<tr>
<td>ko:l</td>
<td>kolu? ✓ ✓ ✓</td>
</tr>
<tr>
<td>liw:a</td>
<td>liwa? ✓ ✓</td>
</tr>
<tr>
<td>hek:a</td>
<td>keka? ✓ ✓</td>
</tr>
<tr>
<td>wiks</td>
<td>wikis ✓</td>
</tr>
<tr>
<td>ho:ja</td>
<td>hoja? ✓ ✓</td>
</tr>
<tr>
<td>hela:j</td>
<td>helaj ✓</td>
</tr>
<tr>
<td>stem</td>
<td>CV: ✓ ✓ ✓</td>
</tr>
<tr>
<td>pul:t</td>
<td>pulu: ✓ ✓ ✓</td>
</tr>
<tr>
<td>hela:j</td>
<td>hela: ✓</td>
</tr>
<tr>
<td>pola:t</td>
<td>pola: ✓ ✓</td>
</tr>
<tr>
<td>hek:a</td>
<td>heka: ✓ ✓</td>
</tr>
</tbody>
</table>

In section 3, it is shown how the ranking of standard faithfulness constraints penalizing these operations predicts the correct interaction of these various phonological processes in the context of the three LH suffixes.

2.2 Syllable structure and stress in SSM

SSM has a length contrast for vowels and for consonants and does not allow complex codas, onsets or adjacent non-identical vowels (Broadbent 1964: 15). Consequently, only the syllable types in (7) are possible in the language. Final consonants are taken to be extrametrical since CVC# syllables count as short and CV:C# syllables are only possible in final position (Freeland 1951: 6).

(7) Syllables in SSM
   a. Short: CV, CVC#
   b. Long: CVC, CVC:, CV:, CV:C#

Syllable weight is crucial for determining stress in the language. Sierra Miwok is an often cited example for iambic lengthening (Callaghan 1987; Hayes 1995; Buckley 1998): Main stress is always on the first heavy syllable and must be on the first or second syllable (Broadbent 1964: 16, 17). From this it follows that any input starting with two light syllables, the second vowel will be lengthened in order to ensure a left-aligned weight-sensitive iamb. The relevant constraints predicting iambic lengthening in my analysis are the standard constraints given in (8a and d) whose effect is exemplified in the tableau in (9) for the abstract input CVCVC. That stress must be on the first or second syllable follows from high-ranked AFL (8a). STRESS-TO-WEIGHT (8b) is the crucial constraint ensuring that only heavy syllables are stressable, excluding the candidates (9a) and (9d). Recall that all non-final coda consonants contribute weight in SSM, I thus simply take MORAICCODA to be high-ranked (8c). Either vowel lengthening or insertion of an epenthetic consonant can
hence make a light syllable heavy. The choice between vowel lengthening in candidate (9c and e) and consonant epenthesis (9g) is decided in favor of the former due to the fact that DepS (8e) penalizing insertion of an epenthetic consonant is ranked above Depμ (8f) penalizing insertion of an epenthetic mora.

(8)  a. **All-Feet-Left (=AFL) (McCarthy and Prince 1993b)**
    Assign a violation mark for every left edge of a foot that is not aligned with the left edge of a prosodic word.

b. **Stress-to-Weight (=StW) (Kager 1999)**
    Assign a violation mark for every stressed syllable that is not heavy (=2μ).

c. **MoraicCoda (=μCoda) (Broselow et al. 1997)**
    Assign a violation mark for every coda consonant not dominated by a mora.

d. **RhyMType:Iamb (=RhT:I) (Kager 1999)**
    Assign a violation mark for every foot with non-final prominence.

e. **DepS (McCarty and Prince 1995)**
    Assign a violation mark for every output segment without an input correspondent.

f. **Depμ (McCarty 2000)**
    Assign a violation mark for every output mora without an input correspondent.

(9) **Iambic Lengthening in SSM**

<table>
<thead>
<tr>
<th>CVCVC</th>
<th>μCoda</th>
<th>AFL</th>
<th>StW</th>
<th>DepS</th>
<th>Depμ</th>
<th>RhT:I</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>(CV.CV)</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>(CV:.CV)</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*!</td>
</tr>
<tr>
<td>c.</td>
<td>(CV.CV)</td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>(CV.CV:)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>e.</td>
<td>CV(CV:)</td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td>(CV.CVʔ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

An additional constraint that is relevant for analyzing syllable structure in SSM is given in (10): superheavy (=trimoraic) syllables are excluded.

(10) **[μμμ]σ (Kager 1999)**
    Assign a violation mark for every syllable associated to three moras.

Another crucial restriction in SSM is the fact that verb stems are maximally bisyllabic. There are various proposals for implementing such maximality restrictions inside OT and I will take it for granted in the following that an Alignment constraint demanding Alignment between a foot and the stem is responsible for this restriction (McCarty and Prince 1993a, and what follows).
I assume in the following that only fully prosodified morphemes and only stems that conform to the the bisyllabicity requirement enter the derivation. This can be ensured under the assumption of Stratal OT (Bermúdez-Otero 2007; Kiparsky 2000; Bermúdez-Otero, Forthcoming) where different derivation steps are evaluated independently with potentially different constraint rankings. In a version of Stratal OT termed “Egalitarian Stratal OT” where it is assumed that “[a]t every stratum, all independent morphological objects undergo phonological evaluation (i.e. all morphological objects which are not part of other morphological objects)” (Trommer 2011: 72). This includes an evaluation of the Lexical Array, hence the set of roots and affixes that are combined to form a word. If now every stem and all affixes are optimized prior to concatenation, it can be ensured that, for example, all vowels are associated with moras and only stems consisting of maximally two syllables become optimal.

3 Analysis for the three LH templates

I argue that the three LH templates in SSM are the simple result of affixing segment-sized phonological structure, namely moras and underspecified segments that are independently argued for in analyses for non-conconcatenative morphology (e.g., Samek-Lodovici 1992; Davis and Ueda 2006b; Haugen and Kennard 2008; Bermúdez-Otero 2012; Bye and Svenonius 2012).

My analysis is based on two simple mechanisms: first, the demand of a first light syllable is predicted from moraic overwriting. A prefixed mora must dominate the first vowel of the stem and is the only mora possible in this syllable. Given the iambic lengthening in SSM, the second syllable is necessarily heavy. This very simple mechanism of moraic overwriting then predicts the class III templates where the second heavy syllable is either consonant- or vowel-final, depending on the number of underlying consonants in the root. The analysis for moraic overwriting in the context of a moraic prefix is presented in section 3.1. Second, the distinctions between final CVC (class I) or CV: (class II) follows from suffixation of defective segments that are minimally specified as consonant or vowel. These radically underspecified segments are realized either as radically underspecified default segments or through fusion with the final segment of the stem. I discuss this in detail in section 3.2.

3.1 Moraic prefixation

The most obvious generalization about the three LH templates is the fact that all consist of a light syllable followed by a heavy syllable. Given the stress system of the language, it is clear that the first part of the generalization is sufficient to describe the prosodic make-up of the templates: that the second syllable is heavy follows from general phonological demands of SSM if the first syllable is light. In this
subsection, I show how this crucial part of all the LH templates is easily predicted from a standard device in phonology, namely affixation of a mora. I assume that in the context of every LH-requiring affix, a mora is prefixed to the root. LH-requiring affixes are consequently circumfixes and consist of a mora that must be realized at the left edge of the stem and a segmental part that is realized at the right edge of the stem. It is therefore taken for granted that every exponent is marked for whether it attaches to the left edge or the right edge of its stem and that circumfixes are split up into two exponents with different requirements for the edge to which they attach, that is, are suffix and prefix at the same time (Spencer 1991; Sproat 1992; Anderson 1992; Marušič 2003).6

That moras exist as (parts of) morphemes triggering lengthening effects is argued for in various analyses of non-concatenative morphology (examples include Lombardi and McCarthy 1991; Samek-Lodovici 1992; Davis and Ueda 2002, 2006ab; Grimes 2002; Flack 2007; Wolf 2007; Saba Kirchner 2007a, 2010; Haugen and Kennard 2008; Yoon 2008; Topintzi 2008; Bye and Svenonius 2012). In contrast to contexts where an affixed mora adds prosodic weight to the base to which it attaches, the prefixed mora in SSM is now assumed to result in overwriting.7 It is integrated into the prosodic structure of the first syllable and makes all further moraic structure in this syllable impossible. This overwriting follows from the constraint (11) that demands that every new association of a segment to a mora must be located at the right edge of a syllable. It is a modified Dep constraint for association lines referring to a specific syllabic position. Since it is a position-sensitive faithfulness constraint, it extends the concept of positional faithfulness (Beckman 1998) to association lines (cf., for example Morén 1999, for DepLinkμ in general). It demands that no epenthetic association line can be added to an underlying association at the left edge of a syllable.

(11) DepLink-μσ (=DL)
Assign a violation mark for every inserted association line between a μ and a segment that is not at the right edge of a syllable.

The effect of DepLink-μσ is illustrated in (13). It derives the output for the stem polat to which a prefixed mora is added.8 As I argue in more detail below, this prefixed μ is the representation I adopt for class III suffixes. In (12), the class III suffix -na is hence part of the input representation as well. Since superheavy syllables are impossible in SSM, affixation of a consonant-initial suffix independently triggers vowel shortening in the second syllable. *[μμμ]σ hence excludes a candidate like (12a).

Owing to the standard markedness constraint *Float (e.g. Wolf 2007; Saba Kirchner 2010), the mora cannot remain unassociated as in candidate (13b). The undominated constraint MaxμAf (12c) demands preservation of every affix mora and deletion of this affixed mora as in candidate (13c) is impossible as well. The affix mora must therefore be integrated into the prosodic structure of the stem it precedes. Since it must be realized at the left edge of the stem, it must
dominate the first vowel. However, association to this first vowel and the resulting lengthening in candidate (13d) is excluded from $\text{DepLink}-\mu\sigma$. The prefixed mora associates to a vowel that is already associated to a mora underlyingly and consequently, the new association is the leftmost association between mora and syllable – a configuration that is penalized by $\text{DepLink}-\mu\sigma$. The overwriting candidate (13e) hence becomes optimal. The affix mora is circled for ease of exposition.

(12) a. $\ast$\textsc{Float} ($=\ast$\textsc{Fl}) (Saba Kirchner 2010)
Assign a violation mark for every $\mu$ in the output that is not prosodically integrated. ($=\text{it is dominated by a syllable node and dominates a segment}$)
b. $\text{max}\mu$ (McCarthy 2000)
Assign a violation mark for every $\mu$ in the input without an output correspondent.
c. $\text{max}\mu\text{Af}$
Assign a violation mark for every affix-$\mu$ in the input without an output correspondent.

(13) Moraic Overwriting

| $\mu\mu\mu\mu$ | $\ast[\mu\mu\mu]\sigma$ | $\ast$\textsc{Fl} | $\text{max}\mu\text{Af}$ | DL | $\text{max}\mu$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. $\mu\mu\mu\mu$</td>
<td>$\ast[\mu\mu\mu]\sigma$</td>
<td>$\ast$\textsc{Fl}</td>
<td>$\text{max}\mu\text{Af}$</td>
<td>DL</td>
<td>$\text{max}\mu$</td>
</tr>
<tr>
<td>b. $\mu\mu\mu\mu$</td>
<td>$\ast[\mu\mu\mu]\sigma$</td>
<td>$\ast$\textsc{Fl}</td>
<td>$\text{max}\mu\text{Af}$</td>
<td>DL</td>
<td>$\text{max}\mu$</td>
</tr>
<tr>
<td>c. $\mu\mu\mu\mu$</td>
<td>$\ast[\mu\mu\mu]\sigma$</td>
<td>$\ast$\textsc{Fl}</td>
<td>$\text{max}\mu\text{Af}$</td>
<td>DL</td>
<td>$\text{max}\mu$</td>
</tr>
<tr>
<td>d. $\mu\mu\mu\mu$</td>
<td>$\ast[\mu\mu\mu]\sigma$</td>
<td>$\ast$\textsc{Fl}</td>
<td>$\text{max}\mu\text{Af}$</td>
<td>DL</td>
<td>$\text{max}\mu$</td>
</tr>
<tr>
<td>e. $\mu\mu\mu\mu$</td>
<td>$\ast[\mu\mu\mu]\sigma$</td>
<td>$\ast$\textsc{Fl}</td>
<td>$\text{max}\mu\text{Af}$</td>
<td>DL</td>
<td>$\text{max}\mu$</td>
</tr>
</tbody>
</table>

It is clear that the moraic overwriting in such a context with a short first syllable does not result in any surface effect for this syllable: the first stem syllable $ko$ was light underlyingly and it is light in the output. However, if the moraic prefix attaches to a stem with an underlyingly heavy first syllable, shortening of this syllable is expected. This is illustrated in the tableau in (14) where the stem $ho:ja$ with a long vowel in the first syllable is optimized. As before, the prefixed mora must dominate the first vowel and it is the only possible mora in the first syllable. That the affix mora is simply added to the moras of the first syllable as in candidate (14b) is again excluded from $\text{DepLink}-\mu\sigma$ (in addition, three-moraic syllables are generally impossible in SSM). Consequently, candidate (14e) apparently
wins the competition and the underlingly long vowel is predicted to be realized as a short vowel.

(14) Moraic overwriting with long first vowel

<table>
<thead>
<tr>
<th>(p) µ µ µ µ µ</th>
<th>#FL</th>
<th>MaxµAf</th>
<th>DL ]</th>
<th>Maxµ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. µ µ µ µ µ</td>
<td>#!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. µ µ µ µ µ</td>
<td>#!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. µ µ µ µ µ</td>
<td>#!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. µ µ µ µ µ</td>
<td>#!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. µ µ µ µ µ</td>
<td>#!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

However, given our knowledge of the stress system of SSM, it is clear that candidate (14c) has no possible grammatical output. A short first syllable necessarily results in a heavy second syllable. The ranking that is responsible for this iambic lengthening was illustrated in (9). Quite parallel to the competition there, the optimal output for the stem µ + hoja is hoja. Vowel lengthening applies to ensure that the second syllable is heavy and hence can be stressed. As I already introduced (cf. (6)), a variety of other phonological operations apply to ensure that the base conforms to the template required by these suffixes. The list in (16) summarizes the effects we find in contexts of class III suffixes, hence in contexts where the initial syllable must be light and the second consequently heavy. It can be seen that metathesis (16C), insertion of an epenthetic vowel (16D and E) and vowel shortening (16F) are triggered from prefixing a mora in addition to vowel lengthening (16A). In several contexts, different strategies could in principle ensure that the second syllable is heavy. In such cases, the ranking of standard faithfulness constraints predicts a preference for certain phonological operations. The relevant ranking arguments mirrored in the application of different phonological operations to make syllables heavy are summarized in (15).

(15) Preference for strategies forming a second heavy syllable

a. V-lengthening is preferred over C-epenthesis: cf. (16A and B)

DepS ≫ Depµ

b. Metathesis is preferred over deletion and V-lengthening/C-epenthesis: cf. (16C)

MaxS ≫ Lin
c. V‐epenthesis and lengthening is preferred over V‐epenthesis and C‐epenthesis:
cf. (16E)
DepS ≫ Depμ

d. Shortening is preferred over deletion: cf. (16 F)
MaxS ≫ Maxμ

The ranking that results from combining these different ranking arguments is given in (16). Candidates excluded by Stress‐to‐Weight, All‐Feet‐Left, or the constraints ensuring proper realization of the moraic prefix (MaxμAf, *Float, and DepLink‐μσ) are omitted from (16) for reasons of space. All the candidates have therefore a light initial syllable. That shortening applies in the last stem wyli:p in (16F) since superheavy CV:C syllables are only possible in final position in SSM. If a stem ending in a CV:C syllable is followed by a suffix that starts with an onset, such a syllable is expected medially: *wy.li:p.pe. Note that I take it for granted that no additional μ is inserted in (16A–C) and (16F): the initial syllable is shortened in these contexts and hence an unassociated base μ remains that becomes the new second μ of the second syllable.

(16) Mora prefixation

<table>
<thead>
<tr>
<th></th>
<th>MaxS</th>
<th>Maxμ</th>
<th>DepS</th>
<th>Lin</th>
<th>Depμ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. ho:ja</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☞ a. (ho.já:).pe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. (ho.jáʔ).pe</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>B. liw:a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☞ a. (li.wá:).pe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. (li.wáʔ).pe</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>C. halki</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. (ha.lí:).pe</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. (ha.líʔ).pe</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>☞ c. (ha.lik).pe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
Let us recap the analysis so far: I argued the prefixation of a μ and moraic over-writing results in a light initial syllable. Given the general prosodic structure of SSM, this straightforwardly predicts that the second syllable must be heavy: the LH template effect. The ranking of standard markedness and faithfulness constraints (16) then predicts the different operations that apply in order to ensure the LH template for class III suffixes. Crucially, the class III suffixes do not specify whether this second syllable has a coda consonant or a long vowel – it only follows that it must be heavy and this is ensured by the strategy that has the best constraint profile with respect to the constraints I list in (16). Class III affixes are hence assumed to be affixes with a suffixing segmental representation and a moraic prefix.

### 3.2 Affixation of underspecified segments

The crucial difference between class III affixes on the one hand and class I and class II affixes on the other hand is the fact that in the latter the second syllable is determined to be either consonant- or vowel-final. In this subsection, I argue that these restrictions are predicted from the affixation of underspecified segments. The affixation of root nodes is another independently motivated mechanism in analyses for non-concatenative morphology and is assumed to predict instances of mutation, reduplication or epenthesis (Bermúdez-Otero 2012; Bye and Svenonius 2012). I assume that the underspecified segments in SSM have a minimal feature specification characterizing them for being either an obstruent/sonorant/glide or a vowel. Only the former sounds are possible final segments preceding the segmental part of

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>D.  wiks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. (wi.kis).pe</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. (wi.ki:).pe</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c. (wi.ki?).pe</td>
<td>*!</td>
<td>**</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E.  ko:l</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. (ko.li:).pe</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>
| b. (ko.li?).pe | **!

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F.  wili:p</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. (wi.lip).pe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. (wi.li:).pe</td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>
a class I affix and only the latter are possible in the context of a class II affix. I assume that the feature [±vocalic] in the definition given in (17) is the binary feature that distinguishes these classes in SSM. Vowels are the only [+vocalic] sounds and obstruents, sonorants, and glides are all specified for [–vocalic].

(17)  \([+\text{vocalic}]\)  
\(=\text{Absence of a narrow constriction among the articulators}\)

The resulting representation for a class I affix is given in (18). The fully specified segments representing the labial voiceless stop \(p\) and the vowel \(e\) are preceded by a segment only specified for [–voc]. (Note that for ease of exposition I omitted a representation for the alternating length of the \(e\) and the prefixing mora that is part of the affix as well.)

(18)  Example: representation for affix class I –peH

\[
\begin{array}{cccc}
+\text{cons} & -\text{cons} & \text{abbreviated as:} \\
-\text{son} & +\text{son} \\
-\text{voc} & -\text{voc} \\
-\text{cont} & +\text{cont} \\
-\text{nas} & -\text{nas} \\
\text{LAB} & \text{DORS} \\
\end{array}
\]

Realization of a segment that is only specified for the feature [±vocalic] violates various markedness constraints demanding full specification, for example the markedness constraint \(\text{HavePlace}\) (19a). I assume that the insertion of epenthetic segmental features is excluded in SSM by undominated \(\text{Dep-F}\) constraints. The preferred option to interpret such a defective root node is therefore fusion with an adjacent segment. This operation violates \(\text{Uniformity}\) (19b) demanding that every output element corresponds only to one input element.

(19)  \(\text{a. HavePlace (}=\text{HvPl})\)  
(Ito and Mester 1993; Padgett 1994)
Assign a violation mark for every segment without a place feature specification.

\(\text{b. Uniformity (}=\text{Unf})\)  
(McCarthy and Prince 1995)
Assign a violation mark for every output element that corresponds to more than one input element.

In addition, the underspecified defective segment can only be realized at the edge of the stem, namely it must form a contiguous string with the fully specified segments that are suffixed to the stem. This is ensured from the \(\text{Contiguity}\) constraint given in (20).\(^{11}\)

(20)  \(\text{O-Contiguity (}=\text{Cnt})\)  
(Landman 2002)
Assign a violation mark for every instance where phonological portions in the output that belong to the same morpheme and form a contiguous string in the input do not form a contiguous string.

(\text{‘No M-internal insertion.’})
(21) gives the derivation for a context where a class I affix is added to a stem and fusion applies in order to realize the underspecified segment. Recall that the prefixed mora predicts that the optimal surface representation is necessarily LH as was already shown in (14). All candidates in (21) have a light first and a heavy second syllable, I hence excluded undominated constraints that ensure the realization of the prefixed mora (cf. (13)). In contrast to the derivations of a class III suffix in (14), however, the nature of the second syllable is now pre-specified: it must either be consonant- or vowel-final. In (21)–(23), the correspondence relation between input and output segment are notated through indices: numbers for the stem and letters for the affix segments in order to ease readability. Elements with multiple indices in the output result from fusion of two input elements.

The illustrating input in (21) is pel:e followed by the class II suffix -j “verbalizer.” As for the affixed mora, a faithfulness constraint specified for affix material now ensures that the segment only specified for [-voc] cannot simply be deleted (MaxSAf), excluding candidate (21a). If the underspecified segment fuses with the penultimate vowel as in candidate (21b), this causes a fatal violation of Cnt. Fusion with the final stem vowel (21c) hence becomes optimal: the structure only induces a violation of Unf but avoids violations of Cnt and HvPl. This is exactly the surface form we encountered for the class III suffix in this context (cf. (14)) – the expected result since this is bimoraic stem.

(21) V-final stem and a class II affix

<table>
<thead>
<tr>
<th>μ+p₁ e₁₂ e₃ e₄ +V• jy</th>
<th>MaxS_Af</th>
<th>Cnt</th>
<th>Id[±v]</th>
<th>HvPl</th>
<th>Unf</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (p₁ e₁₂ l₁ e₄ jy)</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. (p₁ e₁₂ l₁ e₄ jy)</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. (p₁ e₁₂ l₁ e₄ jy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

If, however, this stem is now followed by a class I suffix, additional phonological changes are necessary. In this context, the underspecified segment following the vowel-final base has no chance to fuse with a preceding stem segment and the affix segment remains radically underspecified; it is realized as ?. This derivation is exemplified in (22) where the class II suffix -meH “a person who is … ” is added to the root pel:e. Candidates (22b) and (22c) are possibilities to fuse the underspecified segment with a stem segment. In (22b), the [-voc] root node fuses with another [-voc] segment, namely l. This, however, results in a discontinuous affix string since the stem vowel a intervenes between the two affix segments l₁ and m₁. In candidate (22c), fusion applies between two adjacent segments but results in a fatal violation of Id[±v] since e₄ is specified for [+voc] while c₄ is specified for [-voc]. Consequently, one of these feature specifications must change its value in order to form a licit segment. Since all the fusion strategies to provide a place specification for the underspecified segment fail, it is realized as default coda ? as in (22d).
(22) V-final stem and a class I affix

<table>
<thead>
<tr>
<th></th>
<th>$\mu + p_1 e_2 l_3 e_4 + c \cdot m_5 e_H$</th>
<th>MaxS$_{Af}$</th>
<th>CNT</th>
<th>Id[±v]</th>
<th>HvPL</th>
<th>UNF</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>$(p_1 e_2 l_3 e_4).m_5 e_H$</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>$(p_1 e_2 l_3 x e_4).m_5 e_H$</td>
<td>*!</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>$(p_1 e_2 l_3 \tilde{x} e_4).m_5 e_H$</td>
<td>*!</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>$(p_1 e_2 l_3 \tilde{x} x e_4).m_5 e_H$</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

At this point, another crucial assumption needs to be discussed, namely the possibility that the underspecified segment could simply fuse with the first affix consonant. For (22), this would result in a candidate *(p_1 e_2 l_3 e_4).m_5 e_H that perfectly satisfies O-CONTIGUITY, HAVEPLACE, and IDENT[±voc]. Given that fusion is in fact association of feature bundles, a constraint that prohibits such a configuration is ALTERRATION proposed in van Oostendorp (2007) (cf. for a slightly different version van Oostendorp 2012). The constraint assigns a violation mark for all new (=inserted) association lines between elements affiliated with the same morpheme and easily predicts morphologically derived environment effects that arise since a new association line (~spreading/assimilation) is only possible across morpheme boundaries.

Let us turn to the derivation of consonant-final stems: for class I suffixes, the same surface form is predicted as in the context of a class III affix; and for class II suffixes, additional phonological operations apply to ensure that the segment only specified for [+voc] can fuse with a preceding vowel. In (23), examples are given where the stem hela:j is followed either by the class I suffix ‐kuH (23i) or the class II suffix ‐ équipé (23ii). In (23i), only vowel shortening applies in the winning candidate (23I,d) in order to avoid a superheavy syllable. No additional operation is necessary since the stem already conforms to the LH requirement and the consonantal affix root node can fuse with the stem-final glide j. Fusion with a non-adjacent segment (23i,b) is excluded as well as insertion of an epenthetic consonant (23I,c). In (23ii), however, the underspecified segment cannot simply fuse with the stem-final segment (23b) since this results in a violation of Id[±v]. Fusion with the final vowel in (23c) avoids this violation but is impossible as long as the final consonant intervenes between this vowel and the affix due to O-CONTIGUITY. The optimal candidate is hence (23d) where the stem-final consonant is deleted and the final vowel is faithfully realized as long.

(23) i. C-final stem and a class I suffix

<table>
<thead>
<tr>
<th></th>
<th>$\mu + h_1 e_2 l_3 a_4 j_5 + c \cdot k_u H$</th>
<th>*[μμμ]σ</th>
<th>MaxS$_{Af}$</th>
<th>CNT</th>
<th>Id[±v]</th>
<th>HvPL</th>
<th>UNF</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>$(h_1 e_2 l_3 a_4 j_5).k_u H$</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>$(h_1 e_2 l_3 a_4 j_5).k_u H$</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>$(h_1 e_2 l_3 a_4 j_5).k_u H$</td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>$(h_1 e_2 l_3 a_4 j_5).k_u H$</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ii. C-final stem and a class II suffix

<table>
<thead>
<tr>
<th></th>
<th>μ + h₁e₂l₃a₄j₅ + v₆ty</th>
<th>*μμσ</th>
<th>MaxSAf</th>
<th>Cnt</th>
<th>ID[±v]</th>
<th>HvPL</th>
<th>Unf</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>(h₁e₂l₃a₄j₅t₆)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>(h₁e₂l₃a₄j₅t₆)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>(h₁e₂l₃a₄j₅t₆)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>(h₁e₂l₃a₄j₅t₆)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In contrast to the different derivations for class III suffixes (cf. (16)), the complete constraint ranking penalizing different phonological operations is hardly ever relevant for an analysis of class I and class II affixes. This is simply due to the fact that far fewer possibilities can ensure that a base ends in a consonant or vowel respectively. In most contexts, alternative strategies are simply harmonically bounded since they cause a superset of the violations that the winning strategy causes. An example is a monosyllabic stem like *wɨks that is followed by a class I affix. In principle, deletion and epenthesis could ensure that the stem conforms to an LH template and ends in a consonant: *wɨkɨʔ. It is clear, however, that only epenthesis violates a subset of constraints wiks and is hence the optimal output. Only for CVCCV stems, two different competing strategies could ensure that the base ends in a consonant. The stem ʔalma could be realized as *ʔamaʔ preceding a class I suffix, or as ʔamal. In the former case, deletion and epenthesis applied, and in the latter case, metathesis. That metathesis is preferred over deletion and epenthesis follows from the ranking MaxS ≫ Lin that I already established above (15).

A final important restriction that I already mentioned above is the fact that stems must always be bisyllabic. This excludes abundant epenthesis in order to avoid deletion of segments. In the derivation (23ii), for example, where a stem ending in a consonant precedes a class II suffix, a candidate *hela:j only violates lower-ranked DepS whereas winning hela: violates MaxS. In the former candidate, however, the stem is tri-syllabic.

To summarize this analysis for the three different templates, the representations for the three LH affixes I assume are given in (24). All of them have in common that a mora attaches to the left edge of the stem and results in moraic overwriting as was argued in section 3.1. Class I and class II affixes have an additional radically underspecified segment in their representation that attaches to the right edge of the stem and is either specified for [+voc] or [–voc].

(24) Representations for the three LH affixes

classI: μ + \( + \overset{c}{\text{peH}} \)

classII: μ + \( + \overset{v}{\text{t}} \)

classIII: μ + \( + \overset{}{\text{na}} \)

The affixation of these independently motivated elements (moras, underspecified segments) together with the ranking of faithfulness and markedness constraints I
assumed in (16) and (23) correctly predicts the different phonological operations that apply to ensure that the stems conform to the templatic shapes required by class I–III affixes.

4 Discussion: the nature of the underspecified segments

In my analysis, a segment is a structured featural content that is linked to a root node (=feature geometry Clements 1985; Sagey 1986; McCarthy 1988; Clements and Hume 1995). Most importantly, it is assumed that this root node cannot be empty but minimally consists of the major class features (cf., for example, McCarthy 1988; Selkirk 1991).

Given that the minimal segmental representation is specified for being a consonant or vowel, the model is apparently very similar to the assumption of CV-slots (McCarthy 1979; Marantz 1982; Clements and Keyser 1983). However, there is one interesting aspect in my analysis that distinguishes it fundamentally from classic CV analysis, namely the fact that it assumes affixation of both empty moras and underspecified segments, namely of abstract prosodic timing units and of segmental root nodes. This is actually one of the interesting aspects that Sloan (1991) argues for in her analysis of SSM: the three different LH templates can only be represented properly if one takes into account segmental representations as well as prosodic structure. In my analysis, I adopt this insight and modeled it in a standard OT account.

In the recent literature, moras as well as root nodes are independently motivated in phonological analyses for non-concatenative morphology. An example is Bye and Svenonius (2012) where it is explicitly argued in favor the “[a]ffixation of a root node, possibly specified as a consonant or vowel” (p. 443) in addition to standard mora affixation. They also discuss the interesting fact that the affixation of moras and underspecified segments is often difficult to distinguish on the surface and that it is “far more common to find affixation of featurally deficient root nodes which have some place or manner information” (p. 443). The three LH templates in SSM are now an interesting piece of evidence for exactly such a coexistence of segmental and prosodic template representations.

The present analysis hence is different from the assumptions made in Trommer (Chapter 14, this volume) where segments are radically underspecified (for a similar assumption of “featureless root nodes” cf. Wolf 2007; Saba Kirchner 2007b; Bermúdez-Otero 2012). And it is also slightly different from the predictions made in the alternative analysis in Sloan (1991). There, it is argued that the need to distinguish final CVC and CV:-syllables (class I and II) in contrast to unspecified heavy syllables (class III) in the SSM templates is strong evidence for an analysis assuming (partially) syllabified X-slots (Hyman 1982; Levin 1985; Kaye et al. 1985). The theory hence allows us to distinguish between generic segments, onset segments, nucleus segments and coda segments. This is illustrated with the
representation Sloan (1991) assumes for the three LH templates in SSM in (25). What all three have in common is that they are represented as a light syllable (two X-slots associated as onset and nucleus) that is followed by a heavy syllable containing three X-slots. The difference between class I and class II affixes is the association of the final X-slot: it is associated to the rhyme node (=a coda consonant) or to the nucleus (=a long vowel). The alternating class III templates have a final X-slot that is floating underlyingly. This X-slot is associated on the surface with either the nucleus or the rhyme node, depending on whether a third root consonant is available on the melodic tier or whether all consonants are already associated.


\[
\begin{array}{ccc}
\text{CVCVC} & \text{CVCV} & \text{CVCVX} \\
\sigma & \sigma & \alpha \\
R & R & R \\
N & N & N \\
x & x & x \\
\end{array}
\]

In this framework, it is hence possible to distinguish between consonants, vowels, and unspecified segment nodes in the underlying representation of morphemes. In contrast, the assumption of affix moras and segments being minimally specified for being a consonant or vowel only allows the representation of consonants, vowels, or unspecified segmental length.

5 Conclusion

In this chapter I argued for an analysis of three classes of template-requiring affixes in SSM that relies on the assumption of affixed moras and underspecified segments. I argued that the template-requiring affixes are underlyingly circumfixes: they contain a moraic prefixal part and a segmental suffixal part that might contain radically underspecified segments as well. I showed that the moraic prefix results in moraic overwriting and ensures that the first syllable is necessarily light. The stress system of SSM then predicts that the second syllable must be heavy. Various strategies apply to ensure this for stems preceding a class III suffix where the second stem syllable is either closed or contains a long vowel. Class I and class II, however, demand that the second heavy stem syllable must be either consonant- or vowel-final. This restriction about the nature of the final stem segment follows from the presence of radically underspecified segments in the representation of morphemes. The most interesting conclusion from such an analysis of the three LH templates in SSM is hence the fact that the affixation of two different phonological elements that are in principle segment-sized: moras and underspecified segments.
Acknowledgements

For helpful discussions and comments I would like to thank the audiences of BLS 38 (Berkeley, February 2012), OCP 9 (Berlin, January 2012), the CUNY Conference on the Segment (New York, January 2012), and the colloquium “Neuere Arbeiten zur Grammatiktheorie” (Leipzig, January 2012) where earlier versions of this chapter were presented. I am especially indebted to Jochen Trommer, Ricardo Bermúdez-Otero, and Marc van Oostendorp.

Notes

1 Broadbent (1964) uses some non-standard sound symbols. She uses T for alveolar voiceless stops and y for central high vowels. I replace those with the standard IPA symbols t and ɨ throughout. The dental voiceless stop that is represented as t in Broadbent (1964), is represented as ̪ instead. The symbol Y represents a u if the following syllable contains a u or an o and an ɨ elsewhere. It is the epenthetic default vowel of the language but exists underlyingly as well. The symbol H marks either a preceding long segment, i.e., stands for “;” if it is not followed by another consonant and a juncture or followed/preceded by a C-cluster (except VH + CH). The symbol X represents length as well but in slightly different contexts. It is realized as “;” if a single consonant follows and none precedes the X. Otherwise it is not realized. I follow Broadbent (1964) in using these symbols.

2 They analyze four different stem forms in the Central variety of Sierra Miwok. The fourth stem in Central Sierra Miwok is always CVC.CV, the third stem CVC:VC, and the second stem is either CVCC or CV.CVC. The first stem varies in shape but is restricted through various demands, e.g., the necessity to be bisyllabic and to contain at least one heavy syllable.

3 The forms given in italics are not from Sloan (1991) but are logically concluded from the simple fact that the form preceding a class III affix is identical to the form preceding a class I suffix (for biconsonantal stems) and identical to the form preceding a class II affix (for three-consonantal stems).

4 For discussion cf. for example Chapter 2.3 in Ussishkin (2000) or (Ito et al. 1996).

5 A related concept is TAUTOMORPHEMICITY demanding that morpheme and syllable boundaries should coincide (Crowhurst 1994; Bickel 1998).

6 There are possible arguments for the analysis that the moraic prefix is a morpheme on its own. Freeland (1951) notes that the different templatic forms required by template-requiring affixes can be “regarded as in some measure having grammatical value in itself” (Freeland 1951: 96). All affixes that require the template CVCCV, for example, code the idea of present time, whereas the affixes requiring CVC:VC code the idea of future of past time. If such a semantic decomposition is possible for the LH-requiring affixes as well and if one (abstract) meaning can be found that is common to all LH-requiring affixes, then it would be possible to argue for the existence of an independent morpheme that is represented as moraic prefix. However, such a semantic analysis is rather difficult since the description lacks a detailed discussion of the meaning of affixes. For now, I therefore take it for granted that the LH-template requiring “suffixes” are circumfixes.

Recall the assumption that stems are optimized prior to concatenation. From this it follows that all vowels and non-final coda consonants are moraic in the input. The affix is assumed to be underlyingly mora-less, but nothing hinges on this assumption and the very same result is predicted if a moraic affix (very well possible given the assumption of Richness of the Base) attaches.

This implies that all moras are ordered with respect to each other on the moraic tier, irrespective of whether they are underlingly associated or not.

The details of this assumption and possible alternatives are discussed in section 4.

The definition is slightly modified compared to the original formulation in Landman (2002). O-CONTIGUITY in the version given here only refers to those portions of a morphemes that form a contiguous string in the input. This specification is necessary since class I and class II affixes are assumed to be circumfixes and I took it for granted that the different portions of a morpheme are inherently specified for being realized at the left or right edge of a stem.

I take it for granted that every coda consonant projects its own μ and coda clusters are hence excluded by *[μμμ]σ. Nothing hinges on that implementation and *COMPLEX (Kager 1999) could be the relevant constraint as well. Recall that final consonants are extrametrical, the cluster in (23ii) is hence permissible.

The specification for attaching to the left or right edge is notated by the following/ preceding “+.”

For discussion and literature see, for example, Szigetvári (2011).

References


Subject and Author Index

Abler, W. L., 10, 33.
abstract representations, see empiricism vs. rationalism.
acoustic, basis of element/feature system, 159, 161, 166–167, 171;
distance/similarity, defined, 200;
effects of timing, 79;
level, 4, 7, 10, 14;
affixation, of defective segments, 283–310 passim; of features, 94–95, 284, 303; of mora, 290, 304–308.
allophone/allographic, 49, 200–204, 210; see also contrast; phoneme.
allophonic similarity, 45, 54, 67, 70, 82, 133–140, 157, 222; see also phonetic similarity.
alphabetic writing systems, role in metalinguistic awareness, 8–9, 11–14, 35–36, 40.
Amman, J.C., 2.
Anderson, S.R., 44.
apogee, in fingerspelling, 104–124.
ATR (advanced tongue root), as distinctive feature, 224–233 passim; in element theory, 149, 152, 155, 160–161, 164–165, 256–263; see also tense.
archphoneme, 6; see also defective segments.
Articulatory Phonology, 5, 7, 25–40, 119, 124, 133, 280; see also autosegmental phonology; gesture.
Aslin, R.N., 12–14.
aspiration, 153, 189; phonemic vs. allophonic, 210–212 passim; as secondary feature, 67–68, 77, 97; and VOT, 30, 207–209.
assimilation, in Irish, 92–93; in Korean, 130–131; in Russian, 84, 237, 241–242; in Serbian, 241; across syllables, 91–93; variable, 84; of voicing, 237–242; see also Perceptual Assimilation Model (PAM); spreading.
autosegmental phonology, 44–45, 95, 188, 286–310 passim;  
as precursor to Articulatory Phonology, 7–8.

back (feature), 3, 66, 218–233 passim;  
in element theory, 159, 167–169;  
as secondary feature of consonants, 66–97 passim;  
in Turkish, 257.

Backley, P., 152, 153, 157, 166, 190, 193.

Bell, A.M., 3.

Benjamin, G., 10–11, 18.

Berko, J., 12.

binary, features for defining inventories, 218, 232;  
features in SPE 3, 229;  
as opposed to unary, 149–169 passim;  
see also ternary.

blending systems, 10;  
see also particulate principle.

Bloch, B., 5.

Bloomfield, L., 3, 5, 6.

Bohn, O., 203–204, 254–258 passim, 262;  
Browman, C., 31, 32, 34, 119, 131.

C-center timing, 37.


categorical, as opposed to gradient, 32, 129–133, 135–136, 140, 280;  
perception, 171.

centroid, spectral, 244–246, 248.

chest pulse, 5, 18;  
see also Stetson, H.R.; syllable.


circumfix, affix type, 321, 331.

Cloning Hypothesis, 289–290.

cluster(s), cross-linguistic studies, 202, 203, 205, 248;  
glottal gesture in, 30, 39;  
in Greek, 239;  
in Hungarian, 270;  
in Isthmus Mixe, 95;  
in Japanese, 183;  
in Lithuanian, 92;  
in Marshallese, 93;  
as opposed to single segment, 66–67, 69, 71, 76–85 passim, 219;  
in Päri, 284, 291, 293–310 passim;  
in Russian, 85–91 passim, 241–242;  

coaescence, vs. metathesis in Zoque, 95.

coarticulation, 26–27, 32, 80–81;  
in fingerspelling, 103–124 passim.

coda, and articulatory phasing, 37, 39;  
deletion in Hungarian, 267–268;  
deletion in Spanish, 132;  
as locus of secondary features, 65–71, 76–77, 88–95 passim;  
and mora theory, 279, 292;  
in Päri, 284, 288–292, 303–307 passim;  
see also prosodic categories.

color, as attribute of phonological elements, 149–150, 151, 159, 167–170, 171;  
in Optimality Theory, see Colored Containment Theory under Optimality Theory.

commutation test (for establishing contrast), 5–6.

compensatory action, 28.

compensatory lengthening, 257, 310;  
see also length; mora; weight.

complexity, 26, 71, 85, 155, 169–170, 290.


constraints, in cross-language mapping, 204;  
gradient, 132–133;  
in P-maps, 202;  
on segment inventories, 26;  
morpheme-structure, 290;  
see also Optimality Theory.

constriction, location of, 28, 166.

continuous measure, 132, 136;  
see also categorical; gradient.

contour structure or representation, 180, 188–190, 194.

contrast, as basic concept, 31–32, 46, 59–60, 133–140, 153–171 passim;  
as basis for establishing inventories, 219–232;
as basis for syllable structure, 65–97, 181–183, 191;
and classification of segments, 237–243 passim, 249;
in fingerspelling, 116, 124;
and similarity among segments, 201, 205, 210;
see also contrastive analysis; entropy; inventories; perceptual distinctiveness; phoneme.
Contrastive Analysis Hypothesis, 253–254.
contrastive analysis, of vowel systems, 253–259, 262–263.
corpus, of tokens, for calculating entropy, 137.
coupling relations, 36–38;
within syllables, 37.
coupling strength, 38.
cross-linguistic influence, 203, 205–210, 254.

Darwin, E., 2, 150
defective segments (or roots), 283–284, 297–298, 304, 310, 320, 326;
see also archiphoneme; underspecification; usurpation.
deletion, of association lines, 288;
in Hungarian, 267–268;
in Päri, 297, 302, 308;
in Southern Sierra Miwok, 317, 329;
of segments (or roots), 131.
dependency, in phonology, 151–171, 180–181, 186–196;
see also headedness.
Dependency Phonology, see Government Phonology.
Desensitization Hypothesis, 255, 258, 262;
see also Bohn, O.
devoicing, word final, 9, 26, 32, 202;
diphthongs, in German, 231;
vs. monophthong, 90–91;
vs. triphthong, 219;
see also vowels, long.
discrete, in fingerspelling, 123;
as opposed to continuous, 45, 130, 132, 157;
segments, 2, 5, 8, 10, 27, 33, 200;
see also particulate principle.
discrimination tasks, 201, 205, 244, 253, 258, 261–263;
see also phonetic similarity.
distinctive, see contrast.
Distributed-Morphology, 292.
distribution, complementary, 69, 157;
contrastive vs. allophonic, 134–135, 139;
in cross-linguistic analysis, 201–202, 204, 236–237;
of Greek obstruents, 238–239;
and predictability, 135–136;
of Russian consonants, 241–242;
of segments in Japanese onsets, 183, 186–187;
see also phonotactic constraints.
duality of patterning, 10, 33, 36, 40.
Dunbar, E., 9–10.
durational cues, 253–255.
dynamical systems, 28, 32.
E-language, 25.
Easter egg analogy, 1–2.
electropalatography (EPG) studies, 131.
empiricism vs. rationalism, 3–8.
enhancement, 159.
tropy, 136–139.
epenthesis, in Colored Containment Theory, 286–287;
in Marshallese, 93;
in Southern Sierra Miwok, 319, 323, 329;
as syllable repair, 202.
equifinality, 28.
equivalence classification, in cross-linguistic studies, 211, 254, 262.
event-related potential (ERP) study, 13.
exchange errors, see speech errors.
exemplar models, 9, 133.
extrametrical/extrasyllabic, 71; in Russian, 86–91 passim; in Southern Sierra Miwok, 318. eye-tracking studies, 12, 13.

F1–F2 acoustic space, 162, 223, 257, 271–274; see also formants.

feature bounding, see inventories
feature geometry, 7, 151, 181, 182, 188
Feature Hypothesis, in L2 research, 255, 258, 262.
features, see ATR; back; consonantal; fortis/lenis; front; glottal; high; low; place; round; sonorant; voice.
Firth, J.R., 151.
Flege, J.E., 203–204, 206, 211, 253, 254, 262.
floating elements, 8, 94, 280; in Päri, 290, 292, 303–310 passim.
formant(s), in relation to articulation, 159, 166–167; values in cross-linguistic studies, 257.
fortis/lenis, 51–53, 57, 155; see also length, contrastive.
frequency, as acoustic dimension, 200; effects, 132; fundamental (f0), 159.
front (feature), 3, 159, 220–230 passim; in cross linguistic research, 201, 205; in Japanese, 185–189 passim; as secondary feature of consonants, 66–97 passim.
fusion, in fingerspelling, 122–123, 124; of segments, 320, 326–328.

Gafos, A.I., 32
German as a Foreign Language (GFL), 253, 257, 259, 263.
gestural overlap, 30.
gestures, in Articulatory Phonology, 15, 28–39 passim; as class nodes, 151, 153, 154; phonetic, 28; speech errors involving, 34–35; see also Articulatory Phonology; coupling relations; score, gestural.
ghost segments, 267.
glottal (feature/gesture), 30, 39; in element theory, 153, 155, 160; as secondary feature of consonants, 96.
Goldsmith, J., 7–8, 44–45, 134; see also autosegmental phonology.
Goldstein, L., 31–37 passim, 119, 131, 133.
grain, perceptual scale, 138; see also blending systems; categorical; Kaye, J.

Halle, M., 3, 7, 26, 31, 150–151, 168, 229; handshape, see fingerspelled letters. see also SPE phonology.
heavy, see weight.
high (feature), 3, 224–232 passim; in cross-linguistic research, 201–202, 209; in element theory, 158, 163–164, 169; and glides, 72, 94.
Hockett, C., 1, 2, 5, 10, 33, 150.
Huettig, F.N., 13.
Subject and Author Index

341

I-language, 25.
iambic lengthening, 318–323;
see also length.
Idsardi, W., 9–10.
Information Theory, 136–137, 150.
interface, morphology–phonology, 287;
phonetics–phonology, 10, 203.
interference, in cross-linguistic settings, 207, 254.
International Phonetic Alphabet (IPA), 2, 69, 97, 204, 218, 222, 223, 229.
inventories, feature bounding, 218;
comparisons of to determine similarity, 201–202, 204, 210;
phoneme, 26, 133, 248–249;
see also basis for establishing inventories, under contrast.
Jakobson, R., 150, 151, 166, 237.
Kaye, J., 49, 50, 58, 149–152 passim, 160, 161, 182;
see also Government Phonology.
Kelso, J.A.S., 28, 38.
Kiparsky, P., 50, 159, 229, 320.
Labov, W., 132
Lado, R., 253–254.
lax, see tense.
length, contrastive, 45–60 passim;
in English, 45–46, 49–60 passim, 231;
in Estonian, 45–49, 53–60;
in Italian, 46–47;
in L2 perception, 254–259, 261–263;
in Päri, 284–285, 289–290, 305–310 passim;
representation of, 218;
in Southern Sierra Miwok, 314–323 passim;
see also compensatory lengthening;
iambic lengthening; fortis/lenis; mora; weight.
lexical entry, 7–8, 10–11, 30.
light syllables, see weight.
Lindblom, B., 26.
linguistic experience, in cross-linguistic research, 205–207, 255, 262.
loanword, phonology of, 199, 208–210;
fingerspelling as loanword system, 104.
logographic, 35.
Lombardi, L., 8, 188.
long vowels, constituency of, 218–220, 226, 306;
in cross-linguistic research, 255, 256–257, 258–263 passim;
in English, 46, 49–57 passim;
in Estonian, 47–49, 50–51, 53–58 passim;
in German, 256;
in Hungarian, 275–276, 279–280;
in Italian, 46;
in Southern Sierra Miwok, 314, 317, 325, 331;
in Turkish, 257–259, 262, 263;
see also diphthongs; fortis/lenis; short vowels; tense.
low (feature), 3, 156, 220, 224–230 passim;
in cross-linguistic research, 256–257, 261.
Maddieson, I., 26, 65–91 passim.
Marcus, G.F., 12
margin, as locus for generalizations, 65–97 passim.
McAllister, R., 255, 258, 262.
mentalism, 6, 7, 12–13.
merger of segments, 39, 95, 184, 284;
in sound change, 137.
melting, 9, 12, 14, 35.
metathesis, in Southern Sierra Miwok, 317, 323, 329;
see also coalescence.
minimal pairs, see commutation test; contrast.
Minimality Hypothesis (MH), 50–51;
see also Government Phonology.
mirror neurons, 166.
monophthong, 91, 186, 194, 231;
see also diphthong.
mora, as affix in Päri, 304–308, 310;
as affix in Southern Sierra Miwok, 314–315, 320–331 passim;
floating, 290;
in Japanese, 194;
models of representation based on, 185, 279–280;
as skeletal element, 287.
moraic overwriting, 320, 322–323, 325, 329, 331.
motor theory of speech perception, 166. Munhall, K., 38.
mutation, see consonant mutation.
neighborhood density, as measure of phonological similarity, 199.
Neogrammarians, 2–3.
neutralization, 9, 91–93, 131, 134;
blocking of in cross-linguistic context, 210;
in Päri, 285;
and perception, 138;
in Serbian, 237;
see also assimilation; devoicing.
Non-Arbitrariness Principle (NAP), 49–51;
see also Government Phonology.
non-concatenative morphology, 315, 321, 325, 330;
in Päri, 284, 305, 308–310.
in Southern Sierra Miwok, 320;
Non-derived Environment Principle, 50.
onlinear, affixation, 314.

obligatory contour principle (OCP), 187.
obstruent, see sonorant.
Ohalà, J.J., 4, 8, 238.
onset, and articulatory phasing, 37, 39;
final consonants as, 52;
in Greek, 239;
in Irish, 83–84;
in Japanese, 180–194 passim;
in Lithuanian, 91–92;
as locus of secondary features, 65–71, 76–77, 95;
in Päri, 284, 291, 294, 296;
in Russian, 83–91, 241;
in Southern Sierra Miwok, 318, 324, 330;
see also C-center timing; coda; voice onset time.
Opponent Principle, 149–150, 154–161 passim;
as responsible for binarity, 165.
Optimality Theory (OT), 26, 32, 280;
Colored Containment Theory, 286–287, 310;
stochastic OT, 132;
oscillatory nature of speech, 36.
overlong (vowels in Estonian), see long vowels, in Estonian.
overwriting, moraic, 320–325 passim, 329.
P-map, 202;
in cross-linguistic research, 203–204.
Padgett, J., 91, 241, 249.
Pānani, 2.
partial contrast, 134, 140.
Particle Phonology, 151, 182.
particulate principle, 10, 33.
perception, categorical, 130, 170–171;
in cross-linguistic research, 199–200, 204–211 passim, 253–263 passim;
gradient nature of, 138–139;
motor theory of speech, 166–167;
speech, 6, 26.
Perceptual Assimilation Model (PAM), 205, 253.
perceptual distinctiveness, 26;
see also contrast; inventories; P-map.
perceptual similarity/linkage, cross-linguistic, 199, 202–203, 206–209.
phoneme, vs. allophone, 54;
as bundle of features/components, 3, 5, 182, 201;
as taxonomic entity, 5–6, 31;
as type of segment, 4, 5, 44–45.
phonemic split, 137.
phonetic similarity, 157;
as condition for classifying allophones, 222;
Subject and Author Index

cross-linguistic, 201–203, 243;
gradient scale, 138–139;
see also allophonic similarity;
discrimination tasks.
phonological relationships, gradient vs.
categorical, 133–140 passim;
see also phonological similarity.
phonological similarity, 199;
cross-linguistic, 204–207;
levels of, 200, 210–211;
see also loanword phonology;
neighborhood density; phonetic
similarity.
phonotactic constraints, 199, 205;
in Articulatory Phonology, 33, 34;
in Serbian, 237;
see also distribution.
pinky extension, see fingerspelled letters.
place (features), 6, 82;
assimilation of, 92, 130–131;
in element theory, 151–171 passim;
organization of, 66–70 passim;
see also color, as attributes of
phonological elements.
Port, R., 8, 9, 35.
Posti, L., 46–49 passim.
Pouplier, M., 34.
PRAAT, 244, 259, 269.
Prague school phonology, 3, 6–7.
predictability, gradient, 135–138;
in phonology, 49, 134–135.
Principle of Contrast, 222–225;
see also contrast.
prative, 72, 97, 151, 188.
probabilistic approach, 133–140.
prosodic categories/hierarchy, as locus of
distinctive oppositions, 65–97
passim;
in morphological templates, 316,
320–321, 330–331;
in Optimality Theory, 288–291;
see also coda; onset.
prosody, as fundamental to phonology,
3, 5;
in the lexicon, 12;
see also prosodic categories.
psychological reality, 4, 130;
see also mentalism.
Pycha, A., 308–309.
quantal theory, 166, 221.
rationalism, see empiricism vs. rationalism.
Read, C.Y., 9, 12, 14.
reduplication, 11, 289.
relative prominence (in DP-based
theories), 151–152.
relative phonetics, in acoustic analysis,
245–247;
in cross-linguistic research, 201–202.
relative timing (of articulations/gestures),
30, 37, 66, 78–79, 82, 84–85, 88.
resizing theory, 308–309.
root node, as affix, 285, 292, 297, 300–310
passim, 315, 325–330 passim;
in feature geometry, 96, 185, 188.
as locus of distinctive oppositions,
69–70, 93;
round (feature), 218, 225–232 passim;
in cross-linguistic research, 201–202,
205;
in element theory, 152–155, 167–170;
as enhancer of backness, 159;
as secondary feature of consonants,
66–97 passim;
in Turkish, 257.
Saltzman, E., 37.
Sanders, G., 151.
Schane, S.A., 151.
score, phonetic/phonological/gestural, 7,
29–30, 36–38 passim;
see also gestures.
secondary, articulations 66, 71, 75–97
passim.
seglet, 68–71 passim, 96.
segmental licensing, 66, 69–71, 76–78;
see also prosodic categories, as locus of
distinctive oppositions.
self-diversifying systems, 10, 33.
Sendlmeier, W.F., 258–259.
serial order (of segments), 30.
Shattuck-Hufnagel, S., 34.
short vowels, 46, 48, 226;
in cross-linguistic research, 258–263;
in English, 218–219;
in German, 255–259 passim;
in Hungarian, 268, 271, 277, 279;
in Päri, 284–285, 307;
in Southern Sierra Miwok, 315, 317–318;
see also fortis/lenis; long vowels.
shortening of vowels, 57–58;
in Päri, 305–309 passim;
in Southern Sierra Miwok, 317–328 passim.
Sievers, E., 5.
signed language, 26, 124;
see also fingerspelled letters.
Silverman, D., 8–9.
skeleton, phonological, 8, 310;
in dependency phonology 52;
timing slot (X-slot), 188, 218, 310, 314, 330.
sonorant (feature or segment type),
237–242, 248–249, 325;
in Päri, 283, 293–309 passim;
in Russian, 241–242;
sonority, element, in dependency based phonology, 151, 152, 157;
hierarchy, 39;
sequencing, 85–89 passim;
see also prosodic categories/hierarchy, as locus of distinctive oppositions.
sound change, 2–3;
as gradient phenomenon, 137–138.
Sound Pattern of English (SPE) Phonology, 3, 7, 8, 26, 31, 44, 151, 182, 221, 229.
spectral cues, 253–255, 261, 262.
speech errors, 33–35.
Speech Learning Model (SLM), 206–207, 254, 262.
spreading, of secondary features, 76;
in fingerspelling, 107, 115;
see also assimilation.
Steele, J., 3.
Steriade, D., 202.
Stetson, H.R., 3, 4–6.
Stevens, K., 166.
stress, in Estonian, 48–49;
in Southern Sierra Miwok, 318–320;
see also weight.
Sweet, H., 3, 5.
Swingley, D., 12–14.
syllable, as fundamental unit, 3, 5, 18;
constituents of, 6, 34, 37;
margins, features of, see basis for syllable structure, under contrast.
synergies, 28.
Tauli, V., 48–49.
Taylor, J., 166.
template, 314–331.
temporal relations, 32.
tense, in articulatory theory, 161, 162, 165;
in cross-linguistic research, see long vowels;
properties of, 224;
as property of length, 46, 54, 231;
see also ATR.
ternary, see binary.
tier, of gestural score, 30.
timing, in fingerspelling, 107–108, 115, 124, 133;
phonetic, 37, 38;
triphthong, see diphthong.
Trubetzkoy, N.S., 49, 150–151, 254.
underspecification, 152, 169, 292, 329–330;
in fingerspelling, 124;
in Southern Sierra Miwok, 314–315, 325–329;
see also archiphoneme; defective segments.
usurpation, phonological, 297–299;
see also defective segments.
variable rules, 132.
variation, free, 69, 84;
gradient nature of, 139–140;
see also in fingerspelling under coarticulation.
Verner’s Law, 2–3.
vocal tract, 26–30, 32, 166, 221.
voice onset time (VOT), 130;
in cross-linguistic research, 207–209.
voice/voicing, 50–51;
in cross-linguistic research, 202,
207–210 passim, 219;
in element theory, 152–153;
of fricatives, 236–250 passim;
see also assimilation; devoicing; VOT.
vowels, long see long vowels;
harmony, 26–27, 32, 49–50, 164, 170;
in Turkish, 230;
systems, see inventories;
voiceless, 225–226, 232.

weight, phonological, 48, 279, 305–310
passim, 318, 321;
see also mora; stress.
word-formation rules, 11–14.
writing systems, role in
metalinguistic awareness,
9–14, 35–36, 40.

X-ray, of vocal tract, 28–29.
X-slot, see skeleton.

Yearley, J., 86–87.
Language Index

Abadzakh, 76.
Abkhaz, 73, 77, 98.
Adzera, 73.
Amharic, 74.
Amuzgo, 75.
Ao, 223–224.
Apatani, 223–224.
Arabic, 257, 264;
Moroccan, 32.
Aranda, 73.
Archi, 74, 98.
Awiya, 74.
Axininca Campa, 73.

Beja, 74.
Bella Coola, 74.
Bini (Edo), 73, 98.
Birom, 75.
Breton, 73.
Bulgarian, 74.
Bura, 75.

Caddo, 74.
Chaha, 94.
Chilcotin, 96.
Cofan, 73.

Chinese, 9, 186, 195, 219;
Mandarin, 13, 73, 75, 134, 201–202, 207, 255.
Crow, 82.
Cuna, 74.
Czech, 168, 169.

Dahalo, 74.
Dani, 74.
Diegueno, 74.
Dschang, 73.
Dutch, 61, 201.

English, 5, 12, 30, 33, 45, 46, 49–60 passim, 73, 104, 131, 186, 213, 222–224;
British English, 230–231;
Canadian, 134–135;
in cross-linguistic research, 134, 201–209 passim, 255;
New York City Dialect, 132, 219;
sound change in, 137.

Estonian, 46–49, 53–60 passim, 81, 255.

Fante, 73.
French, 73, 77, 94, 186;
in cross-linguistic research, 205–208, 209.
<table>
<thead>
<tr>
<th>Language</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gà</td>
<td>73</td>
</tr>
<tr>
<td>Gaelic, Scottish</td>
<td>75</td>
</tr>
<tr>
<td>Guambiano</td>
<td>74</td>
</tr>
<tr>
<td>Guarani</td>
<td>74</td>
</tr>
<tr>
<td>Gujarati</td>
<td>139</td>
</tr>
<tr>
<td>Hadza</td>
<td>74</td>
</tr>
<tr>
<td>Haida</td>
<td>74</td>
</tr>
<tr>
<td>Harari</td>
<td>94</td>
</tr>
<tr>
<td>Hausa</td>
<td>74</td>
</tr>
<tr>
<td>Hígi (Kamwe)</td>
<td>81</td>
</tr>
<tr>
<td>Hindi</td>
<td>13, 73</td>
</tr>
<tr>
<td>Hopi</td>
<td>74</td>
</tr>
<tr>
<td>Huasteco</td>
<td>74</td>
</tr>
<tr>
<td>Hungariano</td>
<td>26, 267–280.</td>
</tr>
<tr>
<td>Hupa</td>
<td>79</td>
</tr>
<tr>
<td>Iaai</td>
<td>73</td>
</tr>
<tr>
<td>Igbo</td>
<td>73, 75</td>
</tr>
<tr>
<td>Iranxe</td>
<td>74</td>
</tr>
<tr>
<td>Iraqw</td>
<td>74</td>
</tr>
<tr>
<td>Irish</td>
<td>74, 76, 83–85, 91–92.</td>
</tr>
<tr>
<td>Isthmus Mixe</td>
<td>95</td>
</tr>
<tr>
<td>Italian</td>
<td>46–47, 51, 131, 134, 191.</td>
</tr>
<tr>
<td>Japanese</td>
<td>palatal glide in, Chapter 8.</td>
</tr>
<tr>
<td>English</td>
<td>borrowings in, 209.</td>
</tr>
<tr>
<td>Kabardian</td>
<td>76</td>
</tr>
<tr>
<td>Kam</td>
<td>75</td>
</tr>
<tr>
<td>Kanakuru</td>
<td>73</td>
</tr>
<tr>
<td>Karok</td>
<td>73</td>
</tr>
<tr>
<td>Kashmiri</td>
<td>73, 74, 76.</td>
</tr>
<tr>
<td>Kihehe</td>
<td>73, 94</td>
</tr>
<tr>
<td>Kinande</td>
<td>224</td>
</tr>
<tr>
<td>Klamath</td>
<td>73, 77</td>
</tr>
<tr>
<td>Klao</td>
<td>228</td>
</tr>
<tr>
<td>Kohumono</td>
<td>74</td>
</tr>
<tr>
<td>Kolokuma Ijo</td>
<td>74</td>
</tr>
<tr>
<td>Kom</td>
<td>73, 75</td>
</tr>
<tr>
<td>Konyagi</td>
<td>74</td>
</tr>
<tr>
<td>Korean</td>
<td>73, 130–131, 186, 195; in cross-linguistic research, 207–209.</td>
</tr>
<tr>
<td>Kpelle</td>
<td>74</td>
</tr>
<tr>
<td>Kutep</td>
<td>73, 75</td>
</tr>
<tr>
<td>Kwaio</td>
<td>74</td>
</tr>
<tr>
<td>Kwakwala</td>
<td>74</td>
</tr>
<tr>
<td>Kwoma</td>
<td>74</td>
</tr>
<tr>
<td>Lai</td>
<td>75</td>
</tr>
<tr>
<td>Lak</td>
<td>74</td>
</tr>
<tr>
<td>Lakkia</td>
<td>73, 74, 77.</td>
</tr>
<tr>
<td>Late</td>
<td>75</td>
</tr>
<tr>
<td>Lenakel</td>
<td>74</td>
</tr>
<tr>
<td>Lilooet</td>
<td>73</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>74, 76, 91–92.</td>
</tr>
<tr>
<td>Luganda</td>
<td>73, 75</td>
</tr>
<tr>
<td>Luiseno</td>
<td>74</td>
</tr>
<tr>
<td>Lushootseed</td>
<td>74</td>
</tr>
<tr>
<td>Mandarin</td>
<td>see Chinese.</td>
</tr>
<tr>
<td>Margi</td>
<td>73, 75</td>
</tr>
<tr>
<td>Marshallese</td>
<td>73, 75, 80, 91, 93.</td>
</tr>
<tr>
<td>Mazatec</td>
<td>73, 75</td>
</tr>
<tr>
<td>Mixtec</td>
<td>74</td>
</tr>
<tr>
<td>Moro</td>
<td>227</td>
</tr>
<tr>
<td>Movima</td>
<td>74</td>
</tr>
<tr>
<td>Nahuatl</td>
<td>74</td>
</tr>
<tr>
<td>Nambakaengo</td>
<td>74</td>
</tr>
<tr>
<td>Navajo</td>
<td>74</td>
</tr>
<tr>
<td>Nenets</td>
<td>74</td>
</tr>
<tr>
<td>Ngizim</td>
<td>74</td>
</tr>
<tr>
<td>Nimboran</td>
<td>227, 228</td>
</tr>
<tr>
<td>Nootka</td>
<td>74</td>
</tr>
<tr>
<td>Norwegian</td>
<td>220–221</td>
</tr>
<tr>
<td>Nupe</td>
<td>75</td>
</tr>
<tr>
<td>Nzima</td>
<td>73, 75</td>
</tr>
<tr>
<td>Ocaina</td>
<td>73</td>
</tr>
<tr>
<td>Päri</td>
<td>Chapter 14.</td>
</tr>
<tr>
<td>Paya</td>
<td>74</td>
</tr>
<tr>
<td>Picuris</td>
<td>74</td>
</tr>
<tr>
<td>Pohnpeian</td>
<td>74, 81.</td>
</tr>
<tr>
<td>Polish</td>
<td>73, 74</td>
</tr>
<tr>
<td>Quileute</td>
<td>74</td>
</tr>
</tbody>
</table>
Resigaro, 73.
Russian, 73, 76–80 passim, 83–91, 237–239;
phantomics of /v/ in, 242–249.
Rutul, 74.

Saliba, 74.
Sámi, 73.
Scottish, see Gaelic.
Serbian, 237–241;
phantomics of /v/ in, 242–248.
Shona, 73, 75.
Shuswap, 73, 74.
Siona, 74.
Southern Nambiquara, 74.
Southern Sierra Miwok, Chapter 15.
Spanish, 132, 134, 138, 220;
in cross-linguistic research, 201, 209, 255;
Sui, 74.
Swedish, 168, 231;
in cross-linguistic research, 255.

Taishan, 74.
Tarascan, 74.
Telugu, 73.
Temiar, 10–12.
Temne, 73.
Tera, 74.
Thai, English borrowings in, 210.
Ticuna, 74.

Tikar, 73.
Tlingit, 74.
Toda, 73.
Tonkawa, 74.
Tseshahit, 74.
Tsimshian, 75.
Turkana, 174, 232.
Turkish, 26, 230;
in cross-linguistic research, Chapter 12.
Twi, 73, 75.

Ubykh, 73, 74.
Upper Chehalis, 74.

Wantoat, 74.
Warao, 74.
Western Idoma, 73.
Wichita, 74.
Wiyot, 73, 74.
Woisika, 227.

!Xóô, 230–232.

Yatée Zapotec, 73.
Yawelmani, 316.
Yessan Mayo, 74.
Yupik, 74.

Zoque, 74, 77, 94–95.
Zuni, 74.