Syllable Weight

Phonetics, Phonology, Typology

Matthew Gordon

ROUTLEDGE

Studies in Linguistics

Edited by
Laurence Horn
Yale University

A Routledge Series

Studies in Linguistics

Laurence Horn, *General Editor*

Input-based Phonological Acquisition
Tania S. Zamuner

Vietnamese Tone
A New Analysis
Andrea Hoa Pham

Origins of Predicates
Evidence from Plains Cree
Tomio Hirose

Causes and Consequences of Word Structure
Jennifer Hay

The Phonetics and Phonology of Gutturals
A Case Study from Ju'hoansi
Amanda Miller-Ockhuizen

True to Form
Rising and Falling Declaratives as Questions in English
Christine Gunlogson

Lexical, Pragmatic, and Positional Effects on Prosody in Two Dialects of Croatian and Serbian
An Acoustic Study
Rajka Smiljanic

Lenition and Contrast
The Functional Consequences of Certain Phonetically Conditioned Sound Changes
Naomi Gurevich

The Inflected Infinitive in Romance Languages
Emily Scida

Syntactic Form and Discourse Function in Natural Language Generation
Cassandre Creswell

The Syntax-Information Structure Interface
Evidence from Spanish and English
Eugenia Casielles-Suárez

The Ups and Downs of Child Language
Experimental Studies on Children's Knowledge of Entailment Relationships and Polarity Phenomena
Andrea Gualmini

Category Neutrality
A Type-Logical Investigation
Neal Whitman

Markedness and Faithfulness in Vowel Systems
Viola Miglio

Phonological Augmentation in Prominent Positions
Jennifer L. Smith

Enriched Composition and Inference in the Argument Structure of Chinese
Ren Zhang

Discourse Adjectives
Gina Taranto

Structural Markedness and Syntactic Structure
A Study of Word Order and the Left Periphery in Mexican Spanish
Rodrigo Gutiérrez-Bravo

Negation and Licensing of Negative Polarity Items in Hindi Syntax
Rajesh Kumar

Syllable Weight
Phonetics, Phonology, Typology
Matthew Kelly Gordon

SYLLABLE WEIGHT

Phonetics, Phonology, Typology

Matthew Kelly Gordon

Routledge
New York & London

Routledge
Taylor & Francis Group
270 Madison Avenue
New York, NY 10016

Routledge
Taylor & Francis Group
2 Park Square
Milton Park, Abingdon
Oxon OX14 4RN

Routledge is an imprint of Taylor & Francis Group, an Informa business

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 0-415-97609-X (Hardcover)
International Standard Book Number-13: 978-0-415-97609-1 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Gordon, Matthew Kelly.
Syllable weight : phonetics, phonology, typology / Matthew Kelly Gordon.
p. cm. -- (Studies in linguistics)
Includes bibliographical references and index.
ISBN 0-415-97609-X (alk. paper)
1. Grammar, Comparative and general--Syllable. 2. Phonetics. 3. Language and languages--Variation. 4. Typology (Linguistics) I. Title.

P236.G67 2006
414'.6--dc22 2006025492

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the Routledge Web site at
http://www.routledge-ny.com

Contents

Preface

Phonologists have long known that many prosodic phenomena are sensitive to the inherent "weight" of syllables. For example, Latin preferentially stresses closed syllables and syllables containing long vowels over open syllables containing a short vowel (Allen 1973). Closed syllables and syllables containing long vowels are thus heavy in the Latin stress system. In addition to stress, many other prosodic phenomena have been argued to instantiate weight: poetic metrics, tone, compensatory lengthening, etc.

This thesis explores the idea that syllable weight is driven by phonetic considerations. As a starting point in the investigation, results of an extensive typological survey of syllable weight in approximately 400 languages are presented. This survey suggests that weight is not a property of languages, as predicted by most contemporary theories, but rather is more closely linked to the particular phonological phenomenon involved. It is argued that the primarily process-specific nature of weight is attributed to differences in the phonetic demands imposed by different processes. To illustrate the process-driven nature of syllable weight, focus is on phonetic studies of two weight-sensitive phenomena with divergent phonetic underpinnings: weight-sensitive tone and weight-sensitive stress. Weight-sensitive tone is shown to be guided by the requirement that tonal contrasts be realized on a sufficiently sonorous backdrop to allow for auditory recovery of tonal information. Weight-sensitive stress, on the other hand, is argued to be sensitive to a syllable's auditory loudness, which captures the auditory system's net response to an acoustic stimulus over time. Phonetic considerations are demonstrated to both constrain the range of cross-linguistic variation in weight criteria and also to predict the language specific choice of weight criteria for a given phenomenon.

In addition to being phonetically motivated, it is also shown that the phonology of weight is guided by the requirement that the phonological processes manipulate structurally simple classes of segments. Weight distinctions that are too complex phonologically are avoided, even if they provide a better fit to the phonetic map than other simpler criteria. The result is a compromise between phonetic sensibility and phonological simplicity.

This book is a slightly revised version of a UCLA dissertation completed in 1999. Portions of the material have been reproduced with permission from Cambridge University Press from a chapter "Syllable weight" in the book *Phonetically Based Phonology* (edited by Bruce Hayes, Robert Kirchner, and Donca Steriade, 2004, Cambridge University Press). Material has also been reproduced with permission from the Linguistic Society of America from an article "A phonetically-driven account of syllable weight" appearing in the journal *Language* (vol. 78, 2002, pp. 51–80).

Acknowledgments

There are many people to whom I owe a great debt of gratitude for the inspiration, ideas, time, and support that made completion of this dissertation possible. Although their contributions have greatly improved this work, they should in no way be held responsible for the shortcomings that certainly remain.

First of all, I am deeply indebted to my committee members, Bruce Hayes, Ian Maddieson, Pam Munro, Ralph Sonnenschein, and Donca Steriade, for their abundant intellectual and moral support as I've thought and re-thought the data, the analysis, and the views presented here.

My advisor, Bruce Hayes, has contributed to this dissertation and, more generally, my development as a linguist in more ways than could possibly be apparent in reading what follows. He has always been there for me, contributing crucial insights and encouraging me to approach my work with renewed rigor during moments of mental fatigue or uncertainty. Beyond the actual ideas and advice that he has so generously contributed, Bruce has taught me that being a good linguist involves taking risks. For this valuable lesson and many others, I thank him.

Ian Maddieson also contributed greatly to this work through his meticulous comments and thoughtful advice, particularly in matters phonetic. He has taught me the important skill of looking at data closely and extracting the meaningful patterns.

Pam Munro has been a constant source of encouragement and technical advice in all matters linguistic, always steering me toward new and interesting data (as well as speakers). Thanks to her support, I never lost confidence even in moment of doubts and uncertainty.

Ralph Sonnenschein has always been a welcome and reassuring face in the audience; his comments and questions have helped me to keep in mind the place of linguistics in the broader scientific context.

I have benefited enormously from my discussions with Donca Steriade, who has provided me with clever ideas and insights I could not have possibly had on my own. Although she may not agree with everything contained in this book, her comments and suggestions have made this a far stronger work than it would otherwise have been.

Beyond my committee members, there are numerous other people whose data, comments, and ideas appear in the pages that follow. Thanks to the UCLA faculty members who contributed their time and advice. Thanks to Sun-Ah Jun, Pat Keating, and Peter Ladefoged for their helpful input in phonetic and presentational matters. Also, thanks to Russ Schuh for assistance with the Hausa data and discussion. Other people whom I thank for their contributions include the following (I offer my humble apologies in advance to anyone whom I might have unwittingly overlooked here): Adam Albright, Arto Anttila, Diana Archangeli, Marco Baroni, Taehong Cho, Abby Cohn, Katherine Crosswhite, Susanna Cumming, Jack Dubois, Gorka Elordieta, Edward Flemming, Cecile Fougeron, Ed Garrett, Carol Genetti, Steve Greenberg, Mike Hammond, Larry Hyman, Chai-Shune Hsu, Kathleen Hubbard, Larry Hyman, Sun-Ah Jun, Michael Kenstowicz, Pat Keating, Paul Kiparsky, Robert Kirchner, Jody Kreiman, Paul De Lacy, Peter Ladefoged, Margaret MacEachern, Ian Maddieson, Armin Mester, Pamela Munro, David Perlmutter, Janet Pierrehumbert, Doug Pulleyblank, Sharon Rose, Barry Schein, Dan Silverman, Sumiko Takayanagi, Kim Thomas, Motoko Ueyama, Draga Zec, Jie Zhang, Cheryl Zoll, and audiences at the 19th GLOW Colloquium in Athens, the 23rd meeting of the Berkeley Linguistics Society, the 3rd annual SWOT meeting at UCLA in 1996, and at UCLA, Cornell, MIT, USC, Northwestern, University of Hawaii, UCSB, and UCSD for their helpful feedback and suggestions. A warm thank you is also owed to the speakers who so graciously and patiently contributed the phonetic data contained in this book.

Finally, I wish to thank family members who provided me with much needed emotional support and helped me to never get too lost in my linguistic books. Thanks to my mother and father who have always offered me their unconditional support in matters academic and otherwise and who have always encouraged me to pursue my interests and aspirations and to not let fear of failure stand in my way.

And last but certainly not least, a special debt of gratitude is owed to my wife, Rhonda, for always believing in me, bringing me happiness, and showing me that there's more to life than linguistics.

Chapter One

Introduction

1.0 BACKGROUND

Linguists have long observed that certain phonological phenomena in many languages distinguish between "heavy" and "light" syllables (e.g. Jakobson 1931, Trubetzkoy 1939, Allen 1973, Newman 1972, Hyman 1977, McCarthy 1979a,b, etc.). For example, Latin preferentially stressed closed syllables and syllables containing long vowels over open syllables containing a short vowel (Allen 1973). Closed syllables and syllables containing long vowels were thus heavier in weight than open syllables containing a short vowel in Latin.

While the exact definition of syllable weight is elusive, it may be defined very broadly as that property which differentiates syllables with respect to their prosodic behavior. The difficulty in explicitly defining syllable weight lies in determining which prosodic aspects of language fall under the rubric of weight. As indicated by the Latin stress example above, stress figures prominently among those phenomena considered to involve syllable weight. The domain of weight, however, is not limited to stress. Other phenomena that are potentially sensitive to the weight of syllables include poetic metrics, compensatory lengthening, tone assignment, quantitative aspects of syllable structure, and reduplication. We consider here how these weight sensitive processes instantiate weight. In many languages, only certain types of syllables, the heavier ones, may carry contour tones (Hyman 1985, Duanmu 1994a,b). Weight distinctions are also relevant in many poetic traditions, in which the placement of syllables within the meter is governed by their weight (Hayes 1988). Similarly, many languages have constraints on the minimal size of many classes of words, typically content words. In such languages, words that are subminimal, or not heavy enough, are either disallowed or strongly restricted in their distribution (McCarthy and Prince 1986, 1990, 1995a). Many processes that lengthen or shorten syllables or segments also have been argued to fall under the rubric of weight-based phenomena. For example, long vowels do not occur in closed syllables in many languages, a restriction that has been argued to result from constraints on the maximum weight of the syllable (Steriade 1991, Hayes 1995). Reduplication has also been argued to be a weight sensitive process, because the reduplicant in many languages assumes a certain prosodic shape that appears to conform to some weight standard (McCarthy and Prince 1986, 1990, 1995a). All of these processes superficially have in common that they are sensitive to the phonological weight of syllables.

1.1. FORMAL REPRESENTATIONS OF WEIGHT

As the number of prosodic phenomena argued to instantiate syllable weight has grown, the notion of weight has played an increasingly larger role in phonological theory. In response to the burgeoning role of syllable weight in linguistic theory, phonologists have developed simple yet compelling representations of weight grounded in fundamental concepts such as phonemic length, segment count and sonority.

Of these theories of weight, the two that have gained widest acceptance are skeletal slot models, including CV and X slot models (McCarthy 1979a,b, Steriade 1982, Clements and Keyser 1983, Levin 1985), and moraic models (Hyman 1985, Hayes 1989). The appeal of both of these models is that they assume representations that are projected from independently contrastive properties such as segmental and length distinctions. Units of weight, either skeletal slots (in CV and X slot models) or moras, are assigned to segments. Syllables with a greater number of segments logically receive a greater number of weight units. Similarly, contrasts in segmental length are represented by assigning long segments two weight units, while short segments are associated with one unit of weight. Weight distinctions are thus reducible to differences in the number of units of weight in the syllable. Syllables with a greater number of weight units are "heavier" than syllables with fewer weight units.

The link between syllable weight and the representations designed to model it becomes clearer if we consider the case of Latin, using both moraic and skeletal slot models of weight. Recall from above that Latin preferentially stresses closed syllables and syllables containing long vowels over open syllables containing short vowels. The Latin primary stress rule that demonstrates this weight distinction is as follows: primary stress falls on a heavy penultimate syllable, equivalent to a closed syllable (1a) or one containing a long vowel or diphthong (1b). If the penult is not heavy, stress retracts onto the antepenult (1c) (Allen 1973).

(1) Latin stress
   a. kar'pentum 'carriage'
   b. a'mi:ku:s 'friend'
   c. 'simile 'similar' nom., acc.sg. neuter

1.1.1. Skeletal slot models of weight

First let us consider the representation of Latin weight in a skeletal slot model (McCarthy 1979a,b, Clements and Keyser 1983, Levin 1985); the one presented here is that of Levin (1985). In skeletal slot models, the syllable is divided into constituents. Syllables consist of a nucleus, typically a vowel, which may be preceded by one or more consonants (the syllable onset) and also (in many languages) may be followed by one or more consonants (the coda). Together the nucleus and the coda form a constituent termed the rime (or rhyme). Short segments each project a timing position while long segments project two. In Latin, as in virtually all languages, the onset is ignored for purposes of calculating weight (but see Chapter Four). Only segments (and their associated skeletal slots) that belong to the rime contribute to the weight of a syllable.

The Latin weight distinction has a fairly straightforward representation in this model, as shown in Figure 1.1. (A syllable with both a long vowel and a coda consonant is also heavy, of course, since it contains three timing positions in the rime.)

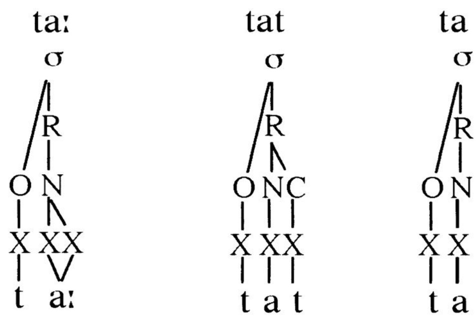

Figure 1.1. Skeletal slot representations of three syllable types

1.1.2. Moraic models of weight

Now let us consider the representation of Latin weight in moraic theory (Hyman 1985, Hayes, 1989). The units of weight in moraic theory are moras.

The weightless nature of onsets is directly captured by assuming that onsets are non-moraic. Contrasts in duration between short and long segments are represented as differences in mora count, parallel to the representation of duration contrasts as differences in the number of timing positions in skeletal slot models. Short segments receive one mora and long segments receive two, as shown in Figure 1.2. In Latin, consonants following a tautosyllabic vowel, i.e. those corresponding to coda consonants in skeletal slot models, are also moraic. The heavy vs. light distinction is thus captured succinctly in terms of mora count; syllables with at least two moras are heavy.

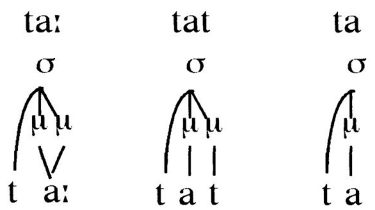

Figure 1.2. Moraic representations of three syllable types

1.1.3. Representations of Weight and Cross-Linguistic Variation in Weight Criteria

Clearly both skeletal slot and moraic models are well equipped to handle weight distinctions of the Latin type according to which closed syllables and syllables containing long vowels are heavy. The Latin weight distinction, however, is not the only weight criterion observed cross-linguistically. Another quite common weight distinction is one that treats only syllables containing long vowels as heavy. For example, in Khalkha Mongolian (Bosson 1964, Walker 1995, 1996), syllables with long vowels, including diphthongs, are heavy, while those containing short vowels are light, whether they are open or not.

The Khalkha weight distinction requires a slight expansion of the principles underlying the representations presented in Figures 1.1 and 1.2. Skeletal slot models must assume that the domain over which weight is calculated may differ between languages. Weight may be calculated over either the entire rime, as in Latin, or over just the nucleus, as in Khalkha. The moraic model must assume that the weight of coda consonants is subject to language specific parameterization, Hayes' (1989) Weight by Position parameter. In languages like Latin, coda consonants are moraic, whereas in languages like Khalkha, they are not.

Interestingly, the Khalkha and Latin type weight distinctions do not exhaust the range of cross-linguistic variation in weight systems. As the database on weight sensitive phenomena available to theoretical phonologists has expanded to include information on a larger cross-section of languages, a diverse array of weight systems has been unearthed, necessitating expansions of the formal apparatus available to theories of weight.

Several languages, e.g. Komi Jaz'va (Itkonen 1955), Chukchi (Skorik 1961, Kenstowicz 1997), Kobon (Davies 1980, Kenstowicz 1997), Yimas (Foley 1991), which base their weight distinctions on neither segment count nor phonemic length contrasts, but rather on vowel quality, have attracted attention in the literature (see Chapter Two). Representing weight contrasts based on vowel quality in terms of differences in the number of weight units is problematic, since moras and skeletal slots are assumed to be projected from contrasts in segment length, not contrasts in segment quality.

Phonologists have also relatively recently noted the existence of languages with greater than binary weight distinctions (see Chapter Two). For example, stress systems in several languages, e.g. Klamath (Barker 1964), Chickasaw (Munro and Willmond 1994), Mam (England 1983, 1986), draw a ternary weight distinction with long vowels and diphthongs (CVV) at the top of the weight hierarchy, closed syllables containing a short vowel (CVC) in the middle, and open syllables containing a short vowel (CV) at the bottom. The representation of a ternary weight distinction of this type as a contrast in numbers of weight units requires that the heaviest syllable in the hierarchy, CVV, receive three moras. This practical necessity, however, violates the principle that representations of weight are projected from contrasts in length. This principle dictates that long vowels should receive two and not three moras. Recent work has even documented the existence of languages with greater than three levels of weight for stress assignment (see Chapter Two), e.g. Kobon (Davies 1980, Kenstowicz 1997), Kara (de Lacy 1997).

1.2. INCONSISTENCY OF WEIGHT CRITERIA

Another standard notion of weight which recent research has shown to be problematic is the view that weight is consistent across phenomena within the same language (Hyman 1985, McCarthy and Prince 1986, 1995b, Zec 1988, Hayes 1989). According to this hypothesis, which I will term the "moraic uniformity hypothesis," all weight sensitive phenomena within a single language observe the same weight criterion and thus employ the same weight representations. Standard representations of weight have captured the assumption that weight is a property of languages by parameterizing weight criteria. For example, Hayes' (1989) moraic theory assumes that coda weight is parameterized; some languages assign a mora to syllable-final (coda) consonants by the Weight by Position parameter, while others do not. Similarly, in skeletal slot models (e.g. Levin 1985), the syllabic affiliation of sonorant consonants is parameterized on a language specific basis: some languages syllabify postvocalic sonorant consonants in the nucleus, while others syllabify them as codas.

Several exceptions to the moraic uniformity hypothesis have surfaced in recent literature, e.g. Steriade (1991), Crowhurst (1991), Hyman (1992), Hayes (1995). For example, Steriade (1991) shows that the stress system, the system of poetic metrics, and the minimal root requirement of Early and Classical Greek are sensitive to different weight criteria from the pitch accent system. At both historical stages of Greek, the stress and metrical systems as well as the minimal root requirement treat both CVV and CVC as heavy. Pitch accent weight criteria are more stringent, however, at both stages. In Early Greek only CVV and syllables closed by a sonorant consonant (CVR) are heavy, while in Ancient Greek only CVV is heavy for purposes of pitch accent placement. A process of vowel shortening in syllables closed by a sonorant also points to the greater weight of CVR relative to syllables closed by an obstruent (CVO) in Early Greek. Crowhurst (1991), Hyman (1992), and Hayes (1995) present additional cases of non-uniformity of weight criteria within a single language.

Cases of conflicted weight criteria are problematic for two reasons. First, they necessarily require reference to at least three levels of weight in a single language. To see this, consider the case of Classical Greek (Steriade 1991). In Classical Greek, CVV is heavy for pitch accent assignment, minimal root requirements, and poetic metrics. CVC is heavy only in the metrical system and for the minimal word requirement but not for pitch accent placement. CV is light for all phenomena. Thus, collapsing all phenomena, CVV is heaviest, followed by CVC, followed by CV. For reasons discussed above, representing this ternary weight distinction is problematic in theories like moraic theory that encode weight distinctions as differences in number of timing positions. A ternary weight distinction requires that the heaviest syllable types, those containing long vowels, carry three moras, but long vowels should only be bimoraic. In fact, the potential for complex weight hierarchies involving more than three levels of weight grows as the number of weight-sensitive phenomena considered increases.

A second challenge presented by cases of conflicted weight criteria concerns the fundamental conception of weight as a language-driven rather than a process-driven phenomenon. Given the increasing number of cases of conflicted weight criteria reported in the literature, it seems worthwhile to explore systematically the alternative and equally plausible hypothesis that weight is more a function of process rather than language. Under this view, variation in weight criteria would be attributed principally to differences between weight-based *phenomena* in the weight distinctions they characteristically employ, rather than to differences between *languages*. For example, it could turn out that weight-sensitive tone tends to observe different weight criteria than weight-sensitive stress and that this process specificity accounts for many cases of conflicted weight criteria. If this scenario turned out to be true, the focus of the theory of weight should shift from explaining how and why languages differ in terms of their weight criteria to addressing how and why weight criteria differ between weight-sensitive phenomena. Exploring weight as not only a language-driven but also a process-driven property also has the potential to provide insight into cases of weight uniformity. To see how examination of the process specific nature of weight is potentially useful, consider the following hypothetical scenario. Let us suppose that coda consonants did not count in determining minimal word requirements in the majority of languages. Similarly, suppose that coda consonants also did not count in determining weight for tone in most languages of the world. This would raise two questions. First, we might ask why codas are characteristically weightless for computing minimal word requirements. Second, we would also want to know why codas are also weightless for purposes of tone in most languages. Crucially, in this hypothetical scenario in which codas are characteristically weightless for both tone and minimal word requirements, even if we were to find a language (in fact, even if we found many such languages) in which coda consonants were weightless for both tone and minimal word requirements, this would not provide support for the view that weight is uniform as a function of language. Rather, assuming that other weight-sensitive phenomena did not display the same cross-linguistic distribution of weight criteria as tone and minimal word requirements, the convergence of weight criteria for tone and minimal words within the same language would be an artifact of the process specificity of weight criteria for these two phenomena: both processes exploit substantially the same weight criterion cross-linguistically. The moral of this story is that, when considering the evidence for uniformity of weight, it is as important to pay attention to the cross-linguistic weight patterns displayed by a single process as to any convergences or divergences of weight criteria within the same language.

Steriade (1991) and Hayes (1995) represent preliminary attempts to introduce process specificity into the theory of weight. These works propose representations designed to account for languages in which one phenomenon or set of phenomena treats both CVV and CVC as heavy while another treats only a subset of these syllable types (usually CVV) as heavy. They also offer tentative observations about which phenomena tend to observe one criterion and which characteristically observe a different criterion. Neither account, however, systematically tests these observations against a large set of data. Furthermore, while both theories attempt to accommodate cases of conflicted weight criteria, neither explicitly addresses the question of whether weight should be modeled as a primarily language-driven or process-driven property.

One of the principal goals of this work is to explore the hypothesis that weight is primarily a property of languages and compare it to the alternative view that weight is mainly a feature of individual processes. To the extent that weight is a property of languages and not of processes, we are justified in parameterizing representations of weight on a language specific basis. If, however, examination of a large set of data demonstrated that weight is more a function of the particular process under consideration than the language involved, we must seek explanations and representations unique to each phenomenon or set of phenomena. Only by examining multiple weight sensitive phenomena in a large number of languages are we able to determine the extent to which weight is determined on a language or process specific basis.

The language- versus process specific nature of weight criteria may be considered from two different angles. First, we may look at languages with more than one weight sensitive process and see whether different phenomena respect the same weight criterion. This is the method that has formed the basis for the discussion up to now. Another possibility is to look at all languages displaying a particular weight sensitive process to determine which criteria are most common for that process cross-linguistically. This can be done for several different weight sensitive processes. Both procedures potentially provide important information about the nature of syllable weight. Weight uniformity will be considered from both of these angles in this work.

Another major goal of this book is to explain the nature and reasons for variation in weight criteria, both variation attributed to process specificity and variation due to language specific properties. We will also propose formal representations of weight and explore how these representations are couched within a formal analysis of weight. The paradigm adopted here for the analyses is that of Optimality Theory (Prince and Smolensky 2004) which is well-suited to capturing the scalar nature of weight and the role of both process specificity and language specificity in the theory of weight.

1.3. THE STRUCTURE OF THE BOOK

The structure of the book is as follows. Chapter Two lays out the basic proposals developed in this work. First, weight will be shown to be more process-specific than language specific. This point will be made based on a typology of weight in approximately 400 languages, focusing on weight-sensitive tone and weight-sensitive stress but also briefly surveying four other weight-sensitive phenomena. Second, Chapter Two advances a proposal that is developed more explicitly in later chapters: that weight criteria are chosen on the basis of a combination of phonetic effectiveness and phonological simplicity. Chapter Two also discusses other assumptions adopted in this book, including the representations of weight employed throughout the work.

Chapters Three and Four present detailed case studies of weight-sensitive tone and stress, respectively, the two phenomena that are the focus of this book. It is in Chapters Three and Four that the grounding of syllable weight in considerations of phonetic effectiveness and phonological simplicity is explored in detail, using phonetic data from several languages. Chapters Three and Four present formal analyses of representative weight-sensitive tone and stress systems, respectively. Chapter Five presents results of the typology of the weight-sensitive phenomena not discussed in detail in Chapter Two, i.e. phenomena other than weight-sensitive tone and stress. Chapter Six provides a summary of the principal findings of the book and directions for future research.

Chapter Two

The Typology of Weight

2.0 A SURVEY OF WEIGHT

This chapter examines the strength of the evidence for weight uniformity both as a function of process and as a function of language. As a starting point in the assessment of the evidence for weight uniformity, a cross-linguistic survey of six phonological phenomena commonly assumed to instantiate weight was conducted. These phenomena are listed in Table 2.1.

Table 2.1. Six weight-sensitive phenomena examined in the survey

- weight-sensitive stress
- weight-sensitive tone
- minimal word requirements
- metrics
- compensatory lengthening
- syllable template phenomena, i.e. closed syllable vowel shortening

Of these six phenomena, the two that serve as the focus of this chapter, and indeed, the focus of the entire book, are weight-sensitive stress and weight-sensitive tone, comparison of which provides crucial insight into the nature of weight as a primarily process-driven phenomenon. Discussion of the other phenomena, which also provides evidence for the process-specific nature of weight, will be limited in this chapter, but will be considered in greater depth in Chapter Five.

2.1. PRELIMINARY TO THE RESULTS: THE SYLLABLE AS A PHONOLOGICAL CONSTITUENT

Implicit in the discussion of weight is the notion that the syllable constitutes a real phonological constituent. If there were no syllables, it would be difficult to characterize the Latin stress rule, since it is the combination of the vowel and the immediately following coda consonant that makes the penultimate syllable heavy, and hence stress-attracting, in Latin (see section 1.1).

Many diagnostics, phonological, phonetic, and psycholinguistic, have been invoked in the literature to determine syllable affiliations, with weight-sensitive processes figuring prominently among the diagnostics. Phonotactic patterns have also been used to decompose words into syllables. For example, if a language does not allow any consonant clusters initially in the word, it is often assumed that a syllable boundary divides clusters of consonants word-internally, the logic being that the language has a uniform constraint against syllable-initial clusters in *all* positions of the word. Similarly, if all words in a language begin with a consonant, the standard conclusion is that any single intervocalic consonant belongs to the same syllable as the following vowel, under the assumption that all syllables must begin with a consonant. Language games have also been used to determine syllable breaks (see Bagemihl 1995 for a survey of the literature on language games). For example, certain languages allow groups of segments to be transposed to another position in the word; thus, the string /pakta/ may become /tapak/ in a hypothetical language game. Segments that move as a single unit, such as /ta/ and /pak/ in the preceding example, are often considered members of the same syllable.

Syllables also appear to have phonetic correlates in many languages. For example, vowels in closed syllables are in most languages phonetically shorter than their counterparts in open syllables (Maddieson 1985). Thus, if a vowel is shortened before a cluster of consonants but not before a single intervocalic consonant, this might provide evidence that at least the first consonant in the cluster belongs to the same syllable as the immediately preceding vowel. Segments may even have different realizations depending on their position in the syllable (e.g. /l/ in many dialects of English); in many cases, these diagnostics accord with other diagnostics of syllable structure. Furthermore, speakers may have strong intuitions about syllable boundaries.

The goal of this book is not to examine the evidence for the syllable as a valid constituent in phonology, an immense topic in and of itself. However, a working definition of the syllable is necessary to discuss phonological weight. For most of the data discussed in this paper, different diagnostics for syllable structure do not conflict with one another. Where there is some question about syllable affiliation, such as occasionally arises in the interpretation of tone bearing units, this will be mentioned in the appropriate places. Unless there is evidence to the contrary in a particular language, I will assume standard syllabification conventions: i.e. a single intervocalic consonant belongs to the same syllable as the immediately following vowel, and a syllable boundary divides sequences of two or more adjacent consonants.

2.2. METHODOLOGY OF THE SURVEY

The primary goal of the weight survey was to examine whether weight tends to be a property of languages or a property of individual phenomena, or a combination of both. As discussed in Chapter One, a large survey of weight can tackle this question from multiple angles. First, we can examine different weight-sensitive processes within the same language to determine the extent to which these processes converge on the same conclusions about weight in a single language. Second, we can look at a particular process in many languages to ascertain whether the same process tends to show the same weight distinctions in many languages. Ultimately, an examination of these issues will enable us to determine the significance of conflicts and convergences in weight criteria.

In order to meet these goals, the survey from which generalizations about weight are drawn must be diverse, both genetically and also, to the extent possible, geographically. Devising a diverse and representative survey is a difficult task for both methodological and practical reasons (see Maddieson 1984 for discussion of many of the issues involved in constructing a typology). The availability of materials on diverse languages is a particularly problematic issue given the type of data sought in a typology of weight. In many cases, consideration of certain weight-sensitive phenomena fell outside the scope of the grammar(s) consulted, through no fault of the authors, since it is quite difficult and time consuming to collect data on weight. In this regard, data on minimal word requirements, metrics, syllable template restrictions, and compensatory lengthening were particularly difficult to locate for most languages. Data on tone and stress, the focus of this work, were easier to access, though descriptions of these prosodic features were often incomplete in the grammars consulted. Generally, the set of grammars that were explicit and thorough enough to enable collection of weight data was only a subset of grammars useful in examining other more transparent phonological properties such as segment inventories.

Given these limitations, the survey must thus include only a subset of languages from each large genetic grouping (termed phyla), in order not to bias the survey in favor of better-documented language phyla. Despite the difficulties inherent in constructing a representative survey of weight, construction of such a survey seems a necessary and worthwhile step in enhancing our understanding of the nature of the weight. Although there will inevitably be gaps or deficiencies in any large-scale typology, I believe that the survey used in this work provides a reasonable gauge for assessing syllable weight on a cross-linguistic basis.

We will now outline the method employed in constructing the typology of syllable weight used in this book. The relatively conservative genetic groupings appearing in the Language Family Index of the twelfth edition of the Ethnologue (Grimes and Grimes 1993) served as the basis for the survey. At least two languages from each of the highest level grouping of languages, which I will term the phylum (i.e. the highest branch in each genetic grouping), were targeted for inclusion in the study. In order to construct a diverse survey, an attempt was made to include no more than two languages from any single language family, the level of classification above the individual languages themselves. As many language isolates and unclassified languages were included as data allowed, since their inclusion did not threaten to bias the data unfairly in favor of certain phyla. Nine language isolates were included in the survey. Only two (Movima and Warembori) of 136 languages that Grimes and Grimes list as unclassified were sufficiently documented to meet criteria for inclusion. (The extreme paucity of data on unclassified languages is perhaps not surprising assuming that languages remain unclassified, as opposed to being classified as language isolates, precisely because they are insufficiently documented to allow for classification.)

The fact that a maximum of two languages per family was targeted as opposed to some other number is the result of a compromise between the goals of achieving as large a survey as possible, but also one that did not unfairly skew the data in favor of better-documented language phyla. For certain language phyla, the target number of languages could either barely be achieved or could not be reached due to a paucity of data. Increasing the number of target languages would thus unfairly bias the survey in favor of better-documented phyla by including more data from better-documented phyla to the exclusion of other less studied phyla. Within each phylum, an attempt was made, as far as resources allowed, to choose diverse languages.

All told 408 languages were included in the survey of weight. Of the 408 languages, 238 are drawn from 19 phyla (each with at least 8 languages represented), with the remaining 159 languages coming from the other 52 phyla (including language isolates and unclassified languages) in Grimes and Grimes. There were 16 small phyla located in either South America or Papua New Guinea that could not be represented in the survey due to a lack of relevant data.

Languages included in the survey of weight appear below in Table 2.2 listed by phylum. Sources consulted for each language appear at the end of the book in Appendix One.

Table 2.2. The genetic classification of languages included in the survey of weight

Phylum	No.	Languages
1. Afro-Asiatic	16	Tamazight, Siwa, Musey, Lamang, Mulwi, Hausa, Somali, Oromo, Iraqw, Dizi, Mocha, Aramaic, Arabic, Tigre, Gurage, Amharic
2. Algic	6	Ojibwa, Menomini, Malecite-Passamaquoddy, Munsee, Blackfoot, Yurok
3. Altaic	10	Buriat, Khalkha, Moghol, Evenki, Even, Turkish, Chuvash, Tatar, Uzbek, Bashkir
4. Andamanese	2	Andamanese, Onge
5. Araucanian	1	Araucanian
6. Arawakan	6	Banawá, Guahibo, Paraujano, Asheninca Campa, Achagua, Arawak
7. Arutani-Sape	0	
8. Australian	6	Maung, Alawa, Djingili, Tiwi, Wardaman, Nyawaygi
9. Austro-Asiatic	13	Khmer, Khmu, Muong, Vietnamese, Sre, Khasi, Pacoh, Brao, Halang, Stieng, Mundari, Santali, Sapuan
10. Austronesian	26	Atayal, Paiwan, Tsou, Kavalan, Larike, Kisar, Kedang, Tetun, Sumbanese, West Tarangan, Loniu, Fijian, Ndumbea, Kara, Patep, Kilivila, Sawai, Murut, Malagasy, Chamorro, Cebuano, Manobo, Karao, Javanese, Malay, Yapese
11. Aymara	2	Aymara, Jaqaru
12. Caddoan	4	Wichita, Pawnee, Kitsai, Caddo
13. Cahuapanan	0	
14. Carib	8	Carib, Hixkaryana, Kashuyana, Pemon, Tiriyo, Waiwai, Machushi, Apalaí
15. Chapacura-Wanham	1	Wari'
16. Chibchan	2	Cuna, Cofán

(continued)

Table 2.2. The genetic classification of languages included in the survey of weight *(continued)*

Phylum	No.	Languages
17. Chimakuan	1	Quileute
18. Chon	1	Selknam
19. Chukotko-Kamchatkan	4	Chukchi, Alutor, Koryak, Kamchadal
20. Coahuiltecan	1	Tonkawa
21. Creole	10	Nubi, Naga, Belizean Creole, Berbice, Krio, Haitian Creole, Korlai, Sango, Torres Strait, Ndyuka
22. Daic	8	Boyao Ai-Cham, Ching (Mak), Khamti, Lao, Shan, Lung Ming Tai, Thai, Saek
23. Dravidian	10	Malto, Koya, Telugu, Malayalam, Tamil, Brahui, Kolami, Kui, Toda, Gonda
24. East Bird's Head	0	
25. East Papuan	2	Anem, Yele
26. Eskimo-Aleut	6	Aleut, Greenlandic, North Alaskan Inupiaq, Central Yupik, Pacific Gulf Yupik, Sirenik
27. Geelvink Bay	0	
28. Gulf	3	Atakapa, Chitimacha, Tunica[1]
29. Hokan	11	Diegueño, Karok, Kashaya Pomo, Eastern Pomo, Mojave, Paipai, Salinan, Yana, Hualapai, Maricopa, Kiliwa
30. Huavean	1	Huave (de San Mateo del Mar)
31. Indo-European	14	Albanian, Gaelic, Icelandic (Old), English, Hindi, Gujarati, Farsi, Maithili, Latin, French, Russian, Czech, Lithuanian, Greek
32. Iroquoian	6	Mohawk, Oneida, Onondaga, Cayuga, Seneca, Cherokee
33. Japanese	3	Japanese, Yonaguni, Amami (Shodon)
34. Jivaroan	1	Jivaro
35. Katukinan	0	

(continued)

Table 2.2.—*(continued)*

Phylum	**No.**	**Languages**
36. Keres	1	Acoma (Western Keres)
37. Khoisan	6	Nama, Kung (Zu|'Hõasi), Sandawe, Gana-Khwe (‖Ani), Naro, !Xõõ
38. Kiowa Tanoan	4	Kiowa, Tiwa, Taos, Jemez
39. Kwontari-Baibai	0	
40. language isolates	11	Ainu, Andoke, Basque, Burushaski, Cayubaba, Gilyak, Korean, Tol (Jicaque), Warao, Yuchi, Zuni
41. Left May	0	
42. Macro-Ge	5	Apinayé, Kaingang, Xavánte, Kayapó, Canela-Krahô
43. Maku	0	
44. Mascoian	0	
45. Mataco-Guaicuru	0	
46. Mayan	8	Aguacatec, Cakchiquel, Chontal, Huasteco, Mam, Tolojolabal, Yucateco, Tsotsil
47. Miao-Yao	1	Green Hmong
48. Misumalpan	0	
49. Mixe-Zoque	4	Sierra Popoluca, Zoque, Totontepec Mixe, Coatlán Mixe
50. Mosetenan	0	
51. Mura	1	Mura-Pirahã
52. Muskogean	2	Chickasaw, Koasati
53. Na Dene	8	Haida, Navajo, Chiricahua Apache, Slavey, Hupa, Tolowa, Sarsi, Tlingit
54. Nambiquaran	2	S. Nambiquara, N. Nambiquara (Mamaindé)
55. Niger-Congo	24	Fula, Wolof, Kisi, Ijo, Jukun, Bushong, Diola, Tunen, Ganda, Bete, Kru, Fong, Anufo, Senoufo, Kabiye, Luganda, Mumuye, Krongo, Moro, Tura, Bobo, Vai, Mende, Kpelle

(continued)

Table 2.2. The genetic classification of languages included in the survey of weight *(continued)*

Phylum	No.	Languages
56. Nilo-Saharan	14	Didinga, Turkana, Shilluk, Nandi, Lango, Fur, Kunama, Runga, Tubu, Songai, Mbai, Mangbetu, Bagirmi, Yulu
57. North Caucasian	4	Kabardian, Lak, Abkhaz, Ubyx
58. Oto-Manguean	6	Mixtec, Otomi, Mazatec, Mitla Zapotec, Trique, Comaltepec Chinantec
59. Paezan	2	Cayapa, Paez
60. Panoan	5	Amahuaca, Capanahua, Cashinahua, Chacobo, Marubu
61. Peba-Yaguan	1	Yagua
62. Penutian	8	Klamath, Sierra Miwok, Nez Perce, Tsimshian, Tzutujil, Wintu, Yawelmani Yokuts, Maidu
63. Quechan	4	Huallaga Quechua, Inga Quechua, Junin-Huanca Quechua, Quicha
64. Salishan	5	Kalispel, Lushootsheed, Comox, Lillooet, Halkomelem
65. Salivan	0	
66. Sepik-Ramu	6	Abau, Alamblak, Boiken, Hewa, Mayo, Yimas
67. Sino-Tibetan	14	Mandarin, Cantonese, Nocte, Tangsa, Tangkhul, Gurung, Dzongkha, Tibetan, Dumi, Burmese, Maru, Lahu, Karen, Nung
68. Siouan	6	Winnebago, Crow, Stoney, Lakota, Mandan, Assiniboine
69. Sko	0	
70. South Caucasian	2	Georgian, Laz
71. Subtiaba-Tlapanec	0	
72. Tacanan	2	Cavineña, Tacana
73. Torricelli	4	Arapesh, Au, Urim, Yil
74. Totonacan	1	Totonac (Misantla)
75. Trans New Guinea	14	Amele, Tauya, Usan, Biangai, Sentani, Abelam, Kobon, Dani, Telefol, Eipomek, Orya, Orokolo, Nankina, Murik

(continued)

Table 2.2.—*(continued)*

Phylum	No.	Languages
76. Tucanoan	4	Barasano, Siona, Siriano, Coreguaje
77. Tupi	5	Guarani, Kamayurá, Júma, Kayabí, Émérillon
78. Unclassified	2	Movima, Warembori
79. Uralic	14	Selkup, Nganasan, Nenets, Hungarian, Ostyak, Komi, Mari, Mordvin, Eastern Sámi, Northern Sámi, Finnish, Estonian, Veps, Votic
80. Uto-Aztecan	8	Hopi, Kawaiisu, Nahuatl, Cora, Tepehuan, Tubatulabal, Luiseño, Comanche
81. Wakashan	2	Kwakw'ala, Nuuchahnulth
82. West Papuan	6	Tidore, Mai Brat, Ternate, Sahu, Tehit, Pagu
83. Witotoan	2	Ocaina, Huitoto
84. Yonamam	1	Sanuma
85. Yenisei	1	Ket
86. Yukaghir	1	Yukaghir
87. Yuki	1	Wappo
88. Zamucoan	0	
89. Zaparoan	1	Arabela

2.3. RESULTS OF THE SURVEY OF WEIGHT-SENSITIVE STRESS AND TONE

In sections 2.3.1–2.3.2, I discuss the results of the typology of weight, focusing on two aspects of weight: the consistency (or lack of consistency, as it turns out) of weight criteria for different phenomena within a single language, and the consistency (or lack thereof) of weight criteria for a single phenomenon in different languages. Examination of these two aspects of weight will provide an answer to the fundamental question of whether weight is primarily a property of languages or of individual processes. We will now turn to discussion of stress and tone. A complete list of all languages in the survey and their weight criteria for all processes examined appears in Appendix Two.

2.3.1. Weight-Sensitive Stress

One of the most commonly invoked diagnostics for syllable weight is weight-sensitive stress (Allen 1973, Hyman 1977, McCarthy 1979a, b). In languages with weight-sensitive stress systems, there are certain syllable types that tend to attract stress based on their relatively greater weight. This differs from weight-*insensitive* languages, which assign stress to a fixed syllable, e.g. the final or penultimate syllable or the stem, regardless of the internal makeup of syllables in the word. In section 1.1, we saw an example of weight-sensitive stress in Latin, in which a CVC(C) or a CVV(C) penult attracts stress. If the penult is neither CVC(C) nor CVV(C), stress retracts to the antepenult.

2.3.1.1. Weight-Sensitive Stress: Results of the Survey

Of the 408 languages in the survey, 314 (or 77.0%) are described as having either stress or pitch accent systems. (This figure does not include languages for which sources did not discuss stress.) Of these 314 documented accent languages, 4 possess pitch accents systems in which words contrast in terms of the presence or absence of a pitch accent. These languages will be excluded from the discussion of stress that follows. Languages that may be characterized as pitch accent languages in sources, e.g. Ainu, Kashaya Pomo, and Koasati, but in which each content word carries at least one accented syllable will, however, be included in the survey figures. Although a pitch peak may be the most salient cue to accent, perhaps the only one, in these languages, their accentual system resembles a traditional stress one in that the accent is culminative, meaning every content word possesses a syllable that is more prominent than others (see Beckman 1986 for discussion of differences between stress and pitch accent systems).

Of the 310 languages with culminative accent systems in the survey, 136 (or 43.9%) have stress systems that are at least partially sensitive to syllable weight.[2] For most of these languages, it is primary stress that is weight-sensitive, although there are several (most of them, Uralic) languages in which weight is involved only in the determination of secondary stress (e.g. Finnish, Hungarian, Veps, Votic, Cayapa, and Yapese). The Turkish weight distinction is employed in only a section of the vocabulary consisting of loan words and proper names (Kaisse 1985, Barker 1989, Kornfilt 1990).

In nine languages in the survey, stress is determined by tone. Six of these languages (Barasano, Bobo, Crow, Haida, Kpelle, and Nubi) are reported to be sensitive only to tone in determining stress placement. These languages are thus not included in the figure of 136 languages with weight-sensitive stress. However, three languages (Ijo, Iraqw, and Krongo) have stress systems that are sensitive to both tone and segmental weight; they are thus treated as weight-sensitive languages. In all languages in which tone is a factor in stress assignment, it is always a high or a contour tone containing a high tone that attracts stress. Bobo makes a three-way weight distinction, with high tones heaviest, followed in turn by mid tones and low tones (Le Bris 1981). A reason for the affinity of stress and high tone will be proposed in Chapter Four.

Of the 136 weight-sensitive languages, 118 employ binary weight distinctions, while 18 are sensitive to complex weight hierarchies involving more than a binary distinction. If we first confine discussion to the binary weight distinctions, certain weight distinctions are far more common than others. If we consider only languages which are diagnostic for differentiating between weight criteria, i.e. languages that both allow coda consonants and have either long vowels or diphthongs, the CVV(C) heavy distinction, which is found in 35 languages, and the CVV(C), CVC heavy distinction, which is found in 42 languages, predominate. This latter figure includes Stoney Dakota, in which CVCC and CVVC are heavy; this distinction has in common with the CVV(C), CVC heavy criterion its treatment of coda consonants as weight-bearing. An additional 5 languages were not assigned to either the CVV(C), CVC heavy or the CVV(C) heavy criteria since they lack coda consonants, thus preventing comparison of the weight of CVV(C) and CVC. Two Yupik languages treat closed syllables as heavy in word-initial syllables, but treat only CVV(C) syllables as heavy elsewhere in the word. Three languages in the survey, Seneca, Oneida, and Kashaya Pomo, treat CVC as heavier than CVV for pitch accent placement. This distinction is unattested in stress systems in the survey.[3]

After the Latin and Khalkha type distinctions, there is a sharp drop-off in the frequency of other weight distinctions for stress. The full vs. reduced vowel distinction is the next most common criterion, but is found in only 12 languages. Note crucially that the term “reduced vowels” in the context of weight refers to *underlying* short central vowels and not to vowels that have undergone post-stress reduction, as for example occurs in English. A variant of this criterion, according to which only reduced vowels in open syllables are light (i.e. all closed syllables and open syllables containing a full vowel are heavy), is found in five languages. Three languages (Kwakw’ala, Nuuchahnulth, and Inga Quechua) treat both long vowels and syllables closed by a sonorant as heavy. Another language, Orya, treats syllables closed by a sonorant preceding another sonorant as heavy. Two languages (Mayo and Yimas) treat low vowels as heavy, while one language (Komi Jaz’va) treats non-high vowels as heavy. In Lamang, full vowels and all syllables closed by a sonorant are heavy, while reduced vowels in open syllables and in syllables closed by an obstruent are light. Kamchadal and Mundari treat syllables containing a glottal stop as heavy. The weight of syllables closed by a glottal stop will be discussed in Chapter Four.

Of the 18 languages with ternary weight distinctions, the most popular type of hierarchy treats long vowels as heaviest, followed by closed syllables containing a short vowel, followed by open syllables containing a short vowel: i.e. CVV(C) > CVC > CV. This weight hierarchy is found in six languages; in one of these languages (Mam) syllables closed by a glottal stop are equivalent in weight to long voweled syllables. The relative frequency of the CVV(C) > CVC > CV hierarchy relative to other ternary weight distinctions is perhaps not surprising, given that it is a conflation of the two most common binary distinctions.

The only other complex weight criterion found in more than one language is the CVVC, (CVCC) > CVV, CVC > CV distinction, which occurs in three languages in the survey. In two of these languages (Pulaar Fula and Eipomek), CVCC syllables do not occur. The only language that has CVCC syllables and employs this three-way distinction, Hindi, has a stress system that is the subject of controversy (see M. Ohala 1977 for discussion).[4] The remaining complex weight distinctions for stress are quite diverse, capitalizing either on differences in vowel quality, vowel length or the presence or absence of coda consonants, or on a combination of these factors.

A list of languages and their weight distinctions for stress appears in Table 2.3. Except where otherwise indicated, the indicated syllable types are those that are heavy. Note the following conventions adopted in Table 2.3 and subsequent tables displaying the distribution of weight criteria. A superscripted [a] indicates that the language does not have codas in any position that is relevant for diagnosing weight for stress. A superscripted [b] indicates an absence of phonemic long vowels. A superscripted [d] indicates an absence of obstruent codas. A superscripted [e] indicates an absence of phonemic long vowels and diphthongs. VV stands for both long vowels, if found in a language, and diphthongs unless otherwise noted. Onset consonants are irrelevant for the processes examined unless indicated otherwise, but are merely included following standard convention. R stands for a sonorant coda, K for a voiceless consonant, G for a voiced consonant. Shaded languages do not contain coda consonants and are thus not useful for diagnosing at least certain weight distinctions.

Certain important generalizations emerge from discussion of these stress systems. First, there are no languages with clear stress accent systems in which consonants are heavier than vowels. The only three languages in Table 2.3 that superficially appear to violate this generalization are Oneida, Seneca, and Kashaya Pomo, all of which possess prominence systems based on pitch accent. In the languages with stress accent systems, if closed syllables containing a short vowel are heavy in a language, then by implication, long vowels (and diphthongs unless phonetically short) will also necessarily be heavy. Crucially,

Table 2.3. Languages with weight-sensitive stress

Language	Phylum	Stress
Aguacatec	Mayan	CVV(C)
Aleut	Eskimo-Aleut	CVV(C)
Aymara	Aymara	CVV(C)
Buriat	Altaic	CVV(C)
Cayuga	Iroquoian	CVV(C)
Cherokee	Iroquoian	CVV(C)
Comanche	Uto-Aztecan	CVV(C)
Fijian[a]	Austronesian	CVV(C)
Huasteco	Mayan	CVV(C)
Hupa	Na Dene	CVV(C)
Ijo[a]	Niger-Congo	CVV(C), H or contour tone
Iraqw	Afro-Asiatic	CVV(C), H tone[5]
Karok	Hokan	CVV(C)
Kawaiisu[a6]	Uto-Aztecan	CVV(C)
Khalkha	Altaic	CVV(C)
Koasati	Muskogean	CVV(C)
Krongo	Niger-Congo	CVV(C), H tone[7]
Kunama	Nilo-Saharan	CVV(C)
Luiseño	Uto-Aztecan	CVV(C)
Malagasy[a]	Austronesian	CVV(C)
Malayalam	Dravidian	CVV(C)
Malecite Passamaquoddy	Algic	CVV(C)
Malto	Dravidian	CVV(C)
Menomini	Algic	CVV(C)
Mojave	Hokan	CVV(C)
Murik	Trans New Guinea	CVV(C)

(continued)

Table 2.3. Languages with weight-sensitive stress *(continued)*

Language	Phylum	Stress
Nganasan	Uralic	CVV(C)
Nyawaygi	Australian	CVV(C)
Ojibwa	Algic	CVV(C)
Quechua, Huallaga	Quechan	CVV(C)
Quechua, Junin-Huanca	Quechan	CVV(C)
Selkup	Uralic	CVV(C)
Telugu	Dravidian	CVV(C)
Tibetan (Lhasa)	Sino-Tibetan	CVV(C)
Tidore[b]	West Papuan	CVV(C)
Tsou[a 8]	Austronesian	CVV(C)
Tubatulabal	Uto-Aztecan	CVV(C)
Winnebago	Siouan	CVV(C)
Wintu	Penutian	CVV(C)
Wolof	Niger-Congo	CVV(C)
Ainu[9]	isolate	CVV(C), CVC
Amele	Trans New Guinea	CVV(C), CVC
Apalaí	Carib	CVV(C), CVC[10]
Arabic	Afro-Asiatic	CVV(C), CVC
Boiken[b]	Sepik-Ramu	CVV(C), CVC[11]
Brahui	Dravidian	CVV(C), CVC
Carib	Carib	CVV(C), CVC
Cayapa	Paezan	CVV(C), CVC except CVʔ[12]
Cebuano	Austronesian	CVV(C), CVC
Cuna	Chibchan	CVV(C), CVC[13]
English	Indo-European	CVV(C), CVC
Estonian	Uralic	CVV(C), CVC[14]

(continued)

Table 2.3.—*(continued)*

Language	**Phylum**	**Stress**
Evenki	Altaic	CVV(C), CVC
Finnish	Uralic	CVV(C), CVC[15]
Greek (Classical)	Indo-European	CVV(C), CVC
Hixkaryana[e]	Carib	CVV(C), CVC
Hopi	Uto-Aztecan	CVV(C), CVC
Huave (de San Mateo del Mar)	Huavean	CVV(C), CVC
Hungarian	Uralic	CVV(C), CVC[16]
Kabardian[e]	North Caucasian	CVV(C), CVC
Kashuyana[e]	Carib	CVV(C), CVC
Khmer	Austro-Asiatic	CVV(C),CVC
Kiriwina[d]	Austronesian	CVV(C), CVC
Korlai[b]	Creole	CVV(C), CVC
Koya	Dravidian	CVV(C), CVC
Latin	Indo-European	CVV(C), CVC
Macushi	Carib	CVV(C), CVC
Maidu	Penutian	CVV(C), CVC
Maung[e]	Australian	CVV(C), CVC
Miwok, Sierra	Penutian	CVV(C), CVC
Mixe, Coatlán	Mixe-Zoque	CVV(C), CVC[17]
Munsee	Algic	CVV(C), CVC
Nambiquara, N.	Nambiquaran	CVV(C), CVC
Ndyuka[d]	Creole	CVV(C), CVC
Nez Perce	Penutian	CVV(C), CVC
Stoney Dakota[e][18]	Siouan	CVVC, CVC
Tepehuan	Uto-Aztecan	CVV(C), CVC
Tol (Jicaque)[e]	isolate	CVV(C), CVC

(continued)

Table 2.3. Languages with weight-sensitive stress *(continued)*		
Language	**Phylum**	**Stress**
Totonac (Misantla)	Totonacan	CVV(C), CVC[19]
Turkish	Altaic	CVV(C), CVC[20]
Veps[b 21]	Uralic	CVV(C), CVC
Votic	Uralic	CVV(C), CVC
West Tarangan[e]	Austronesian	CVV(C), CVC
Yana	Hokan	CVV(C), CVC
Yupik, Central	Eskimo-Aleut	CVV(C); initial CVC
Yupik, Pacific Gulf	Eskimo-Aleut	CVV(C); initial CVC
Kwakw'ala	Wakashan	CVV(C), CVR
Nootka	Wakashan	CVV(C), CVR
Quechua, Inga[e]	Quechan	CVV(C), CVR
Komi (Jaz'va)[e]	Uralic	Non-high V
Mayo[b]	Sepik-Ramu	Low V
Yimas[b]	Sepik-Ramu	Low V
Mundari[b]	Austro-Asiatic	CVʔ(C)
Orya	Trans New Guinea	CVV(C), CVR.R[22]
Kamchadal[e]	Chukotko-Kamchatkan	ʔV, Vʔ
Au	Torricelli	Full V
Chuvash[e]	Altaic	Full V
Javanese[e]	Austronesian	Full V
Lillooet[e]	Salish	FullV[23]
Lushootseed[e]	Salishan	Full V
Mordvin[e]	Uralic	Full V
Nankina[e]	Trans New Guinea	Full V, onset[24]
Ostyak (Vach)[e]	Uralic	Full V
Patep[b]	Austronesian	Full V
Sawai[e]	Austronesian	Full V[25]

(continued)

Table 2.3.—*(continued)*

Language	**Phylum**	**Stress**
Urim[b][26]	Torricelli	Full V
Yil[b][27]	Torricelli	Full V
Aljutor[e]	Chukotko-Kamchatkan	Red V in open syll light
Malay[e]	Austronesian	Red V in open syll light
Mari[b] (literary)	Uralic	Red V in open syll light
Sarangani Manobo[e]	Austronesian	Red V in open syll light
Sentani[e]	Trans New Guinea	Red V in open syll light
Lamang[b][28]	Afro-Asiatic	Full V, CVR
Abelam[b]	Trans New Guinea	CaV > CʌV > Cij, Cuw > CV
Asheninca Campa	Arawakan	CVV > Ca(C), Ce(C), Co(C), CiC > Ci > Cɨ[29][30]
Chickasaw	Muskogean	CVV(C) > CVC > CV
Chukchi[e]	Chukotko-Kamchatkan	Low V, Mid V > High V > Red V
Eipomek[b]	Trans New Guinea	CVVC > CVV, CVC > CV
Fula(Pulaar)	Niger-Congo	CVVC > CVV, CVC > CV
Gujarati	Indo-European	LowV >CəC > Non-low V >Cə
Hindi	Indo-European	CVCC, CVVC > CVV, CVC > CV[31]
Irish (Munster)	Indo-European	CVV(C) > Cax > CV[32]
Júma[e]	Tupi	final syll unless onsetless
Kara	Austronesian	CV: > CaV, CaC > Ca > CV_iV_k, CVC > CNon-low V[33]
Klamath	Penutian	CVV(C) > CVC > CV

(continued)

Table 2.3. Languages with weight-sensitive stress *(continued)*

Language	**Phylum**	**Stress**
Kobon[a,b][34]	Trans New Guinea	Low V> MidV > High V > Red V
Maithili	Indo-European	CVV(C) > CVC > CV
Mam	Mayan	CVV, CVʔ > CVC > CV
Moro[b]	Niger-Congo	CVC > Full V> Red V
Mura-Pirahã[a]	Mura	KVV > GVV > VV > KV > GV
Tamil	Dravidian	CVV(C) > CVC > CV
Yapese	Austronesian	CVV(C) > CVC > CV

the converse is not true; there are many languages such as Khalkha, in which long vowels are heavy but closed syllables containing a short vowel are light. A few languages make a distinction between sonorant and obstruent codas. This distinction also respects an implicational relationship: If obstruent codas are heavy, then sonorant codas will also be heavy. If we conflate the distinction between sonorant and obstruent codas with the distinction between long vowels and closed syllables, an implicational hierarchy emerges with long vowels and diphthongs at the top of the weight hierarchy, followed in turn by syllables closed by a sonorant, syllables closed by an obstruent, and open syllables containing a short vowel, i.e. CVV(C) > CVR > CVO > CV.

There is a second implicational hierarchy for weight in stress systems. Lower vowels are universally at least as heavy as higher vowels. Thus, in languages that make a weight distinction between vowels of different heights, it is always the lower vowels that are heavier than the higher ones. An exception to this generalization is provided by reduced vowels, which although articulatorily lower than high vowels, are nevertheless treated as lighter than high vowels in several languages, notably those in which they are phonetically quite short (see section 4.1.1). In fact, reduced vowels are never heavier than full vowels in any language, suggesting that reduced vowels can be incorporated into the hierarchy of weight based on vowel quality, at the bottom of the scale. The resulting scale is thus Low Vowels > Mid Vowels > High Vowels > Reduced Vowels, where the full vowels are the low, mid, and high vowels. As this hierarchy predicts, although there are no languages which treat reduced vowels as heavier than any of the full vowels, there are languages which treat reduced vowels as lighter than the full vowels.

A second important generalization is that, in virtually all languages in the survey with four[35] exceptions (Júma, Pirahã, Banawá, which are spoken in South America but which are genetically unrelated, and Nankina, a language belonging to the Trans New Guinea phylum), syllable onsets do not play any role in the calculation of weight for weight-sensitive stress. The weightless status of onsets is not a property unique to weight-sensitive stress systems; onsets are characteristically weightless for all weight-sensitive phenomena (Hyman 1985, Hayes 1989). Reasons for this weight asymmetry between onsets and the rest of the syllable will not be explored here. Hopefully further research, perhaps along similar lines to the approach developed here for rime-sensitive syllable weight, will shed more light on the question of why onsets are weightless.[36] Chapter Four contains some discussion of the motivations behind onset sensitivity in stress systems.

2.3.2. Weight-Sensitive Tone

Weight has also been linked in the literature to tonal phenomena (e.g. Woo 1969, Hyman 1985, Zec 1988, Hyman and Ngunga 1994, Duanmu 1994a,b), most commonly to account for languages in which only a subset of syllable types may carry contour tones or rising or falling pitch accents. Henceforth, I will frequently use the term "tone" as a blanket term for both tone and pitch accent phenomena. To take an example of a language with weight-sensitive tone, consider the case of Navajo (Sapir and Hoijer 1967), which tolerates contour tones only on long vowels (CVV)(2a). Syllables containing short vowels may not carry a contour tone, whether they are open or closed (2b).

(2) Navajo tone
a. **kâ:**kì: 'crow', sì**zě:**tí 'my cousin'
b. (hypothetical) ***kân**kì: or ***kâ**kì:

Depending on the language, the asymmetrical ability of some but not other syllable types to carry contour tones may manifest itself in a few different ways, including for example, tone sandhi or spreading phenomena and lexical restrictions against contour tones or pitch accents on certain syllables.

It is typically assumed that contour tones result from the combination of two level tones (e.g. Woo 1969, Hyman 1985, Duanmu 1994a,b). Thus, a rising tone reflects the combination of a low tone followed by a high tone, while a falling tone is represented as a combination of a high tone followed by a low tone. Given the compositionality of contour tones, restrictions against contour tones are often assumed to arise from a prohibition against associations between more than one tone and a single timing position (either a skeletal slot or mora). Thus, because a contour tone consists of two tones, it requires two timing positions on which to be realized, one for each element of the contour, in languages with weight-sensitive tone. For example, the Navajo restriction against contour tones on CVO and CV follows if we assume that only vowels are associated with weight bearing timing positions in Navajo. In moraic representations, this is captured by assuming that only vocoids are moraic in Navajo. Sample representations of a long vowel carrying a level tone and a long vowel carrying a contour tone appear in Figure 2.1.

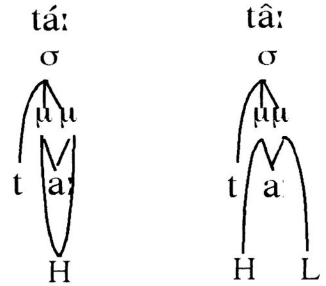

Figure 2.1. Moraic representations of weight-sensitive tone

In a skeletal slot model, we may assume that the relevant weight-bearing constituent for Navajo tone is the nucleus. Branching nuclei, those with two skeletal slots, may support contour tones, as shown in Figure 2.2; each element of the contour links to a timing position associated with a vowel. Syllables that do not contain branching nuclei may not support a contour tone, since tone units cannot be associated with coda consonants and more than one tone may not link to a single timing position.

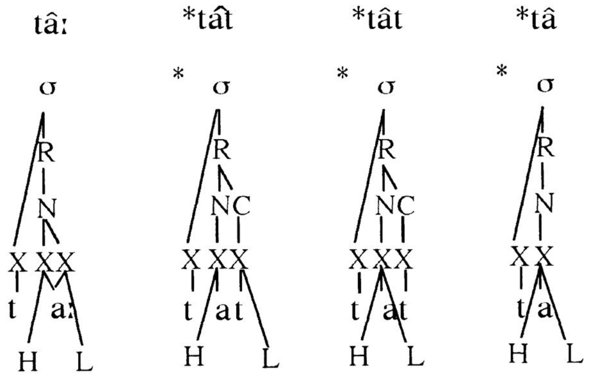

Figure 2.2. Skeletal slot representations of weight-sensitive tone

There are certain complicating issues that arise in the discussion of tonal restrictions of this sort. First, many languages have independent restrictions on syllable structure that preclude evaluation of the weight status of certain syllable types. For example, there are several languages in the survey (e.g. Ijo, Jukun), in which only long vowels may bear contour tones, but coda consonants are not permitted. Though such languages are weight-sensitive, they do not shed light on the question of whether codas are weighted or not. Furthermore and along similar lines, there are many languages in the survey (e.g. Mbai, Tura, Kabiye) that only tolerate sonorant codas. While these languages provide insight into the overall weight status of codas relative to vowels, they offer no evidence, either confirming or disconfirming, for a weight distinction between sonorant and obstruent consonants.

There is another more subtle confounding issue that is relevant to the study of weight. Many authors treat tone as a property of the syllable rather than as a property of segments. For example, if a language only allows contour tones on long vowels or sequences of vowels, one possible interpretation, in fact one commonly adopted, is that all syllables consist maximally of a single short vowel and that the domain of tone is the syllable. Under these assumptions, no syllable may carry a contour tone.

In many cases, this is in fact the correct diachronic analysis: long vowels or sequences of vowels have arisen from disyllabic sequences through loss of the intervocalic consonant. However, this analysis rests on the crucial assumption that a phonetically long vowel on the surface is actually interrupted by a syllable boundary. In most cases, the most compelling synchronic evidence for this interpretation comes from tonal assignment itself, a circular argument; works typically describe long vowels as phonetically single vowels which are neither rearticulated in the middle of their production nor are characterized by changes in amplitude or phonation type, all potential phonetic cues to syllable boundaries. In such cases, I assume for purposes of the survey that long vowels are tautosyllabic, unless there is independent distributional evidence that a syllable boundary divides the long vowel in half.

A related but somewhat less problematic issue concerns the syllabic status of tone bearing nasal consonants in many languages, most notably Niger-Congo languages. In many languages, the only segments other than vowels that may carry tone are nasals. In languages with this characteristic, there is typically independent distributional evidence that nasals are syllabic and pattern with vowels. Some distributional respects in which syllabic nasals may pattern with vowels are as follows. They may head a syllable or they may immediately follow or precede a consonant even if consonant clusters are otherwise prohibited. Furthermore, syllabic nasals may have greater energy and be longer than their non-syllabic nasal counterparts (see Price 1980 on phonetic correlates of syllabic consonants in English). The upshot of this discussion is that the ability of syllabic nasals to carry tone to the exclusion of other consonants often cannot be reliably treated as a property of the nasal as opposed to a property related to their status as syllabic consonants.

2.3.2.1. Weight-Sensitive Tone: Results of the Survey

Of the 408 languages in the survey, 110 (or 27.0%) are described as using tone contrastively at the lexical level or for morphological purposes. Of these 111 languages, 61 have contour tone restrictions that are relevant to the issue of syllable weight. In most of the languages surveyed with weight related restrictions on contour tones, the restrictions operate on the surface. However, in at least one language (Shilluk), the restriction on contour tones holds only at the lexical level and may be violated on the surface.

The remaining 50 tone languages that do not have weight-sensitive tone restrictions either allow contour tones on short vowels in open syllables in addition to other syllable types or do not tolerate tautosyllabic contours at all. Thirteen languages appear to lack contour tones completely. An additional language, Khmu, lacks contour tones, but nevertheless makes a tonal weight distinction: the only syllables preceding the root that may carry tones are those containing a sonorant. Khmu is thus included among those with weight-sensitive tone. This leaves 37 languages that allow contour tones on all syllable types. A large number of these languages (17) have an impoverished syllable structure, either lacking coda consonants completely or limiting them to sonorants or to glottal stop. In fact, languages which possess a rich inventory of syllable types, including open and closed syllables and a wide variety of coda types, all of which can support contour tones, appear to be a rarity cross-linguistically, though they are attested, for example, in languages such as Mangbetu, Mocha, Kunama, and Mulwi. Tolerance of contour tones on all syllable types seems to be most prevalent in Afro-Asiatic and Nilo-Saharan languages.

The two types of syllables that cross-linguistically are most likely to tolerate contour tones are long vowels and syllables closed by a sonorant. A total of 28 languages allow contour tones only on long vowels. Of these 28, four lack coda consonants, and are thus not probative for diagnosing coda weight. A large number of languages, 30, allow contour tones only on long vowels and syllables closed by a sonorant. Of these 30, five have only sonorant codas and are thus not instructive for diagnosing the relative weight of obstruents and sonorants. This leaves 25 that display the CVV(C), CVR heavy criterion for tone. In one of these, Acoma, the sonorant consonant that contributes