UNIVERSITY OF CALIFORNIA

Los Angeles

Gradient Weight in Phonology

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Linguistics

by

Kevin Michael Ryan

2011

The dissertation of Kevin Michael Ryan is approved.

Robert Theodore Daland

Russell G. Schuh

Donka Minkova Stockwell

Bruce P. Hayes, Committee Co-Chair

Kie Ross Zuraw, Committee Co-Chair

University of California, Los Angeles

2011

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ASCII	American Standard Code for Information Interchange
b	binary number suffix
C	consonant
C_1	one or more consonants
CM	coda maximization
dB	decibels
e	Euler’s number (≈2.71828)
H	heavy syllable
HG	Harmonic Grammar
IPA	International Phonetic Alphabet (IPA 1999)
L	light syllable
ms	milliseconds
OM	onset maximization
OT	Optimality Theory
S	strong metrical position
V	vowel
V̆	short vowel
Vː	long vowel
VV	long vowel or (heavy) diphthong
W	weak metrical position
X	single syllable of any weight
.	syllable break
#	word boundary
’	(in scansion) elided or resolved vowel
—	heavy syllable (same as ‘H’)
˘	light syllable (same as ‘L’)
•	caesura
*	ungrammatical
<...>	orthography

ACKNOWLEDGMENTS

First, I wish to express my gratitude to Bruce Hayes and Kie Zuraw, whom I have been fortunate to have as advisers throughout my years at UCLA. Bruce has been instrumental in giving direction to my research in both phonology and morphology, always pushing me towards bigger ideas and more rigorous models. He is a real scientist, whose broad view of the field and eagerness towards new methodologies and technologies will always inspire me. Kie has had a massive impact on almost all of my research at UCLA. As an adviser and as a linguist, she is thorough, dedicated, and above all incredibly insightful. It seems no matter how incoherently I might begin to express an idea to her, a picture would quickly materialize on the whiteboard summarizing it impeccably, accompanied by spot-on follow-ups. As I start out as a professor, I will aspire to Bruce's and Kie's model.

I am also grateful to my committee, Robert Daland, Donka Minkova, and Russ Schuh, for their contributions to this dissertation and other aspects of my research and career. Robert kindly helped with, among other things, experimental materials and modeling issues. I was always flattered by the interest that Donka showed in my work, both for this dissertation and for various prior projects, where her support has been encouraging and her feedback invaluable. Russ has also been a great support over the last couple years for my aspirations in both metrics and morphology, and has always generously reached out to share his expertise and field data.

Numerous other people have contributed to this dissertation and related work in various ways and degrees. I must mention my colleague, collaborator, and friend Dieter Gunkel, who has spent many hours helping me to prepare corpora and to address issues concerning the Indo-European languages. This topic had been on my back-burner for many years and it was in talking to Dieter that I initially decided to return to it to pursue further. Invaluable suggestions along the way also came from Joe Pater, Michael Becker, Arto Anttila, Paul Kiparsky, Kristján Árnason, Ingvar Lofstedt, audiences at UCLA, USC, UMass Amherst, Harvard, Stanford, LSA 2011, WCCFL 2010, anonymous reviewers of a related article, and surely many others I am forgetting. Finally, the National Science Foundation critically supported my early graduate career with a graduate research fellowship.

Though they were not directly involved with this research, I owe a debt of gratitude to my many language and linguistics professors over the years, including my Berkeley advisers Andrew Garrett and George Hart, who both went out of their ways to take me under their wings and who inspired me to pursue linguistics in graduate school. Classes at Berkeley with Ian Maddieson, Sharon Inkelas, and Larry Hyman also left lasting impressions of the excitement and rigor of the field. I am honored to have had Paul Kiparsky as an adviser during my year at Stanford. He is largely responsible, for one, for fostering my initial interest in poetics; in fact, it was at Stanford that I had the idea to investigate syllable weight as a gradient phenomenon. His perspicacity and breadth continue to be an inspiration. At UCLA, Pam Munro, though not involved with my dissertation, was an adviser for other projects and I thank her for her warmth, support, and always useful feedback.

Of course, nothing I have accomplished would have been possible without the love and encouragement of my family, especially my parents, sister, aunt Boppie (whose name I derived at a very young age from /kæθi/), and uncle Bernie. Finally, my fondest memories of graduate school involve my fellow graduate students, great friends and colleagues whom I cannot imagine having done without. More than anyone else, Laura McPherson has been an unceasing source of personal support and crazy schemes over the past few years.

VITA

November 10, 1979	Born, Milwaukee, Wisconsin
2003	B.A. with Honors, Linguistics B.A. with Honors, South and Southeast Asian Studies University of California, Berkeley Berkeley, California
2003	Departmental Citation in Linguistics, Highest Distinction in Letters and Science, Phi Beta Kappa University of California, Berkeley
2004–2005	Doctoral Student in Linguistics Stanford University Stanford, California
2004–2007	National Science Foundation Graduate Fellowship
2005	Linguistic Society of America Institute Fellowship Cambridge, Massachusetts
2005–2010	Research Assistant Phonetics Laboratory and Department of Linguistics University of California, Los Angeles
2007–2010	Teaching Assistant, Associate, Fellow; Instructor University of California, Los Angeles
2007	M.A., Linguistics University of California, Los Angeles Los Angeles, California
2007	Linguistic Society of America Institute Fellowship Stanford, California
2010–2011	Dissertation Year Fellowship University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Ryan, Kevin (2005). Rimeless rhymes. Paper presented at Poetics Fest: Workshop on Language & Poetic Form. Stanford University.

Zuraw, Kie with Kevin Ryan (2007). Frequency influences on phonological rule application within and across words. Paper presented at the Workshop on Variation, Gradience, and Frequency in Phonology. Stanford University.

Ryan, Kevin (2007). The dissociation of melodic and structural correspondence in half-rhyme. Paper presented at Poetics Fest II. University of California, Santa Cruz.

Ryan, Kevin (2008). Gradient morphotactics in Tagalog. Paper presented at the Berkeley Workshop on Affix Ordering. Berkeley, California.

Ryan, Kevin (2009). Released glottal stop and prosodic constituency in Matatlán Zapotec. Paper presented at the Annual Meeting of the Society for the Study of Indigenous Languages of America. San Francisco, California.

Ryan, Kevin (2009). Morphotactic extension: A learning-theoretic explanation of free variation in affix order. Paper presented at the 83rd Annual Meeting of the Linguistic Society of America. San Francisco, California.

Ryan, Kevin (2010). Subsegmental syllabification. Paper presented at the 28th West Coast Conference on Formal Linguistics. University of Southern California.

Gunkel, Dieter and Kevin Ryan (2010). Phonotactic factors in word order in the Ṛg-Veda. Paper presented at the 22nd Annual UCLA Indo-European Conference. Los Angeles, California.

Ryan, Kevin (2010). Contextual and non-contextual prosodic minimality. Paper presented at the 41st Meeting of the North Eastern Linguistics Society. Philadelphia, Pennsylvania. To appear in Lena Fainleib, Nicholas LaCara, and Yangsook Park, eds. *NELS 41: Proceedings of the 41st Meeting of the North Eastern Linguistics Society.* University of Massachusetts, Amherst: GLSA Publications.

Ryan, Kevin M. (2010). Variable affix order: Grammar and learning. *Language* 86.4: 758–791.

Ryan, Kevin (2011). Gradient syllable weight in the tragic trimeter and Homeric hexameter. Paper presented at the 142nd Annual American Philological Association Meeting. *A new look at Greek prosody* panel. San Antonio, Texas.

Ryan, Kevin (2011). A non-sonority-driven consonant weight distinction in Tamil. Paper presented at the 85th Annual Meeting of the Linguistic Society of America. Pittsburgh, Pennsylvania.

Ryan, Kevin M. (2011). Subsegmental syllabification. In Mary Byram Washburn, Katherine McKinney-Bock, Erika Varis, Ann Sawyer, and Barbara Tomaszewicz, eds. *WCCFL 28: Proceedings of the 28th West Coast Conference on Formal Linguistics*. Somerville, Massachusetts: Cascadilla Press. 178–185.

Gunkel, Dieter and Kevin Ryan (2011). Hiatus avoidance and metrification in the Rigveda. In Stephanie W. Jamison, H. Craig Melchert, and Brent Vine (eds.). *Proceedings of the 22nd Annual UCLA Indo-European Conference*. Bremen: Hempen.

Ryan, Kevin (2011). Subcategorical syllable weight. Invited colloquium presented at Stanford University.

Ryan, Kevin (2011). Gradient weight in phonology. Invited colloquium presented at the University of Massachusetts, Amherst and Harvard University.

ABSTRACT OF THE DISSERTATION

Gradient Weight in Phonology

by

Kevin Michael Ryan

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2011

Professor Bruce P. Hayes, Co-Chair

Professor Kie Ross Zuraw, Co-Chair

Research on syllable weight in generative phonology has focused almost exclusively on systems in which weight is treated as an ordinal hierarchy of clearly delineated categories (e.g. light and heavy). As I discuss, canonical weight-sensitive phenomena in phonology, including quantitative meter and quantity-sensitive stress, can also treat weight as a gradient interval scale in which (1) differences between syllable types are matters of relative degree rather than strict domination, and (2) there is no clear segregation of syllable types into categories, but rather a continuous distribution of types along a continuum of weight. In a meter sensitive to gradient weight, progressively heavier syllables are progressively more skewed towards metrically strong positions, all else being equal. Gradient weight is likewise evident in a stress system when syllables vary along a continuum in their propensities to attract stress, again controlling for distributional confounds unrelated to weight.

The dissertation consists of four parts. Part I comprises corpus studies of six quantitative meters, namely, Kambaṉ's Tamil epic meter, the Homeric Greek hexameter, the Latin hexameter, the Finnish Kalevala meter, the Epic Sanskrit *śloka*, and the Old Norse *dróttkvætt*. All six are widely held to treat syllable weight as exclusively binary. I demonstrate that, in addition to distinguishing light and heavy syllables, the poets in all six traditions exhibit sensitivity to a continuum of weight within the heavies, to the extent that I am able to derive some of the most detailed scales of syllable weight yet documented for individual languages (e.g. at least nine levels in the first case study). Moreover, across the six languages, the weight scales are strongly correlated, both with each other and with the crosslinguistic typology of weight-sensitive phenomena, supporting and shedding new light on the universal phonology of syllable weight.

Part II first addresses the universal principles of weight, e.g. complexity and sonority, that motivate the features of the scales in part I. Particular emphasis is given to Tamil, including its violation of the sonority principle: $\mathrm{C_0\breve{V}R}$ (R = rhotic) is lighter than all other $\mathrm{C_0\breve{V}C_{\neq R}}$, despite the rhotics being highly sonorous. Prosodic minimality in Tamil also diagnoses $\mathrm{C_0\breve{V}R}$ as being lighter than $\mathrm{C_0\breve{V}C_{\neq R}}$. I argue that this discrepancy is motivated by the short durations of rhotics relative to other codas in Tamil. More generally, judging from a phonetic corpus of Tamil, duration of the rime (or energy integrated over duration) correlates tightly with the weight continuum inferred from meter. Building on these empirical findings, a generative analysis of gradient weight mapping is proposed in a maximum entropy constraint framework. In it, categorical and gradient constraints (the latter being violated to real-valued degrees supplied by the phonetics; Flemming 2001) interact to generate the weight mapping typology. This typology includes fully categorical systems, fully gradient systems (directly reflecting the phonetics), and systems exhibiting various degrees of incomplete categorization, in which the phonology is polarized towards categories but remains sensitive to the gradient phonetic interface of weight within categories.

Part III treats the contributions of onsets to syllable weight. While onset structure is irrelevant to weight categorization in all the languages examined, it contributes consistently to weight as a statistical effect, in that more complexity is associated with greater weight (e.g. in Old Norse, onset $\varnothing < C < CC < CCC_1$). Even in Tamil, in which complex onsets are illicit, mean duration of the onset correlates significantly with metrical weight.

Finally, part IV considers gradient weight in stress assignment in English. The distribution of stress in extant disyllables follows the same (universal) principles established for meter: First, the complexity of the coda correlates monotonically with stress propensity, such that $\varnothing < C < CC$ $(<)$ CCC_1 (as seen in both nouns and verbs). Vowel length is also important, such that, taking the rime as a whole, the hierarchy $\breve{V} < \breve{V}C < VV < VVC$ is observed for both nouns and verbs. As in meter, onset complexity also contributes significantly to stress propensity, as observed independently in nouns and verbs as well as in initial and final position in disyllables. Experimental evidence is presented supporting the productivity of the universal $\breve{V} < \breve{V}C < VV < VVC$ hierarchy in English stress, as well as the onset effect.

Part I

Gradient syllable weight in quantitative metrics

1 Introduction

In quantitative meter, rhythm is instantiated through mapping conventions regulating the distribution of syllable weight in verse constituents (e.g. Halle and Keyser 1971, Halle 1970, Hayes 1988). Most typically, a distinction between light and heavy syllables is observed, such that certain contexts permit one or the other, but not both. For example, the line-initial position of the Latin hexameter can be filled only by a heavy syllable, as figure 1 illustrates with (a) the opening line of Vergil's *Aeneid* and (b) a hypothetical grammatical but unmetrical comparison.[1]

(a)	ar.ma	.wi.rum.kʷe	.ka.noː	.troː.jaj	.kʷiː	.priː.mu.s	a.b	oː.riːs
(b)	*e.go	.wi.rum.kʷe	.ka.noː	.troː.jaj	.kʷiː	.priː.mu.s	a.b	oː.riːs

Figure 1: A weight restriction in quantitative meter.

Through corpus studies of six metrical traditions, I demonstrate that the poets' manipulation of phonological material in every one is influenced by sensitivity to additional contrasts in syllable weight. These contrasts emerge not as categorical restrictions, but as significant preferences, even while controlling for possible lexical and contextual confounds using mixed effects regression models or Monte Carlo observed vs. expected models.

In particular, I examine the meter of Kamban̲'s Middle Tamil epic, the Homeric

[1] On exceptionality, see §3.1 and fn. 80.

Greek and Classical Latin hexameters, the Finnish Kalevala meter, the Sanskrit epic *šloka*, and the Old Norse skaldic *dróttkvætt*. These corpora range from approximately ten thousand to two hundred thousand lines each. In each corpus, asymmetries in the metrical distributions of syllable types permit the derivation of an interval scale of subcategorical (e.g. intra-heavy) weight. For example, the following skeletal hierarchy recurs across the corpora: $C_0\breve{V} < C_0\breve{V}C < C_0VV < C_0VVC$.[2] Pursued further, this method reveals some of the most articulated syllable weight scales documented for individual languages (see, for instance, the conclusion of the Tamil study in §2), all in meters in which weight is usually assumed to be exclusively binary.

That these scales reflect weight is supported by typological parallels with other weight-sensitive systems, including stress. First, the structure of the rime takes precedence over that of the onset. Second, more structure (e.g. timing slots) correlates with greater weight. Third, even when complexity is held constant, greater sonority is associated with (if anything) greater weight (e.g. $C_0\breve{V}T < C_0\breve{V}N$; though see §7.2 for a caveat concerning light rhotic-final syllables in Tamil). As another example, among stress systems distinguishing the weights of $C_0\breve{V}C$ and C_0VV, the former is almost always the lighter; $C_0\breve{V}C < C_0VV$ likewise holds of every meter examined here. Finally, if one defines a heavy syllable in meter as one that is required/preferred in strong metrical positions (or avoided in weak ones), it is sensible to speak of syllables that are progressively more favored in strong over weak positions (all else being equal) as being progressively heavier.

In sum, although light ($C_0\breve{V}$) vs. heavy is a prominent weight distinction in all six traditions, rising to the level of categoricality in at least some of them, I show that

---

[2] C_0VV and C_0VVC are not significantly different from each other according to the test/corpus used for Sanskrit in §5. Furthermore, a subset of C_0VVC patterns as anomalously light in Latin in §3.6, perhaps owing to closed-syllable shortening.

speakers of these languages were sensitive to various additional contrasts in syllable weight as factors influencing their choices in quantitative versification, a highly conventionalized language game in which syllable weight is manipulated to effect rhythm. Individual languages are like microcosms of the crosslinguistic typology in the gradient realm, in that factors in weight that are ignored for categorization emerge as statistical preferences. These findings are significant for the phonology of weight, for metrical grammar, and for modeling the interaction of categoricity and gradience in the treatment of scalar phenomena (on this last point, see §10).

2 Tamil: Kambaṉ's epic meter

2.1 Metrical and corpus preliminaries

I demonstrate in this section that weight mapping in Tamil meter is underlain by a scale of syllable weight that is considerably more fine-grained than the traditional heavy/light distinction. At least nine grades of weight based on the structure and features of the rime are shown to be significant (see also §12.1 for additional effects concerning the duration of the onset). Moreover, some rime types, such as V̆R (where V̆ is a short vowel and R a rhotic), are intermediate between heavy and light and not clearly affiliated with either binary category in metrical weight mapping.

As a metrical corpus of Tamil, I employ Kambaṉ's[3] *Irāmāyaṇam*[4] epic (critical edition 1956), a Tamil telling of the South Asian Rāmāyaṇa epic (Hart and Heifetz 1988). Kambaṉ lived c. 1200 CE and composed in early Middle Tamil, the standard

---

[3]Variants of this name include Kambar, Kampar, and Kampaṉ. Under the present romanization of names and citations — the most widely employed — *n* is dental, *ṉ* alveolar, and *ṇ* retroflex.

[4]Variants include *Rāmāyaṇam*, *Rāmāyaṇa*, *Irāmāvatāram*, and *Rāmāvatāram*.

medieval (but still largely accessible to present-day speakers) literary dialect. A Unicode Tamil-script version of the poem was obtained from the Tamil Electronic Library (`tamilelibrary.org`, accessed June 2009), converted to a lossless ASCII romanization, and groomed (e.g. by applying sandhi rules, including some discussed by Rajam 1992 and Lehmann 1994) to render the transcription more phonetically transparent. The resulting text comprises 42,128 lines in 10,532 AAAA-rhyming quatrains.

Tamil meter, like all the meters examined in this dissertation, is quantitative, which is to say that the distribution of syllable weight is regulated in verse constituents (e.g. Halle and Keyser 1971, Halle 1970, Kiparsky 1977, Hayes 1988, Hanson and Kiparsky 1996; see §10). The description of Tamil syllable weight and syllabification in the remainder of this paragraph can be considered a standard traditional account (Niklas 1988, Zvelebil 1989, Rajam 1992, Murugan 2000); some aspects of this account will be revisited in the following pages. First, weight in Tamil is claimed to be binary, such that $\mathrm{C_0\breve{V}}$ ($\mathrm{C_0}$ = zero or more consonants; $\mathrm{\breve{V}}$ = short vowel) is light and all other syllables are heavy. As for syllabification, onsets are maximized, but complex onsets are illicit; thus, as is typically assumed: V.(C)V, VC.CV, and VCC.CV. As in many quantitative meters (e.g. Sanskrit), word boundaries are ignored for basic scansion. For example, the first syllable in $\mathrm{C_0\breve{V}C\#V}$ is treated as light.[5] Finally, diphthongs can be treated as V(ː)C sequences, C being the appropriate glide, [j] or [ʋ]. As such, they always scan as heavy. There are, however, two conditional exceptions to this rule, namely, [ăj] and [ăʋ] (as I denote them here). These short diphthongs (which, despite their transliteration, scan as $\mathrm{\breve{V}}$, not $\mathrm{\breve{V}C}$) are

[5]See Ryan (forthcoming) on resyllabification in Tamil.

the realizations of /aj/ and /aʋ/ in any non-initial position.[6]

Kamban̲'s epic comprises an indefinite variety of meters. To begin with two examples, the first and last couplets (also known as distichs or half-verses) of the text (§1a-b and §10,532c-d) are given in figure 2, first in Tamil script, then in IPA transcription (International Phonetic Association 1999), and finally in terms of syllable weight (H = heavy, L = light). Hyphenated word-final segments are the result of gemination across word boundaries (e.g. [pinăj-p paka...]). Weight templates are spaced at word boundaries. Bullet (•) indicates caesura.

§1ab	உலகம் யாவய்யுந் தாம் உள வாக்கலும் நிலை பெறுத்தலு நீக்கலு நீங்கலா
	ulakam jaːʋăjjun̪ t̪aːm uɭa ʋaːkkalum n̪ilăj petut̪t̪alʉ n̪iːkkalʉ n̪iːŋkalaː
	LLH HLH • H LL HLH LL LHLL • HLL HLH
§10,532cd	பராபரம் ஆகி நிந்த பஞ்பினைப் பகருவார்கள் நராபதி யாகி பின்னு நமனய்யும் வெல்லுவாரே
	paɾaːpaɾam aːki ninta paɲpinăjp pakaɾuʋaːɾkaɭ n̪aɾaːpati jaːki pinnʉ n̪amanăjjum ʋelluʋaːɾeː
	LHLL HL HL • HLH LLLHH LHLL HL HL • LLLH HLHH

Figure 2: Scansion of the first and last couplets of Kamban̲'s epic.

[6]In addition to shortening, [ăj] is often monophthongized to a mid vowel by modern speakers. /aʋ/, for its part, is rare. The exclusively word-initial heaviness of these diphthongs might be motivated by initial-syllable privilege (e.g. Beckman 1998 on Tamil), accent (often claimed to be word-initial in Tamil, e.g. Keane 2003, 2006, Krishnamurti 2003; cf. Kiparsky 2003 on the monomoraicity of unstressed diphthongs in Finnish), and/or prosodic minimality, in the following sense. Tamil observes a strict bimoraic minimum on prosodic words (Ryan forthcoming). Light diphthongs might be therefore be coerced into bimoraicity, so to speak, when uttered as the rimes of monosyllables (cf. Morén 1999, Blumenfeld 2010). When light diphthongs are initial in polysyllables, they are almost always derived from monosyllabic roots. Thus, even when minimality is not at stake, it might still exert an influence through analogy or enforcement of minimality prior to suffixation.

In these examples, syllable count per line matches within the two couplets but not between them. Furthermore, syllables tend to correspond in weight between lines of the same verse, as indicated here by underlining, but across the two verses, the sequences have little in common.

2.2 On the metrical diversity of Kamban̲'s text

This section provides some more background on the metrical composition of Kamban̲'s epic. I do not attempt to provide an adequate description (much less generative analysis) of the meter(s), as such a (substantial) undertaking is unnecessary for the present goal of describing the treatment of syllable weight. As already hinted in §2.1, there is at first glance no single metrical template — or even small number of metrical templates — underlying the text.[7] While some overarching metrical tendencies may well exist (e.g. medial caesura, periodicity, etc.), the meters are superficially quite diverse (cf. Deo 2007 on the diversity of Classical Sanskrit meters). At the same time, however, there is no doubt that the text is quantitatively regulated, with tight correspondences (at least within quatrains) in syllable count, syllable weight, and word boundary distribution, as the examples in figures 2 and 5 suggest.

The distribution of line length (in terms of syllable count) in Kamban̲'s epic is characterized as a histogram in figure 3 and as a longitudinal plot (over the course of the 10,532 quatrains) in figure 4. Eleven to sixteen syllables is most typical, sixteen being the mode (each x-axis tick refers to the bar to its left), but the distribution

[7] This situation might be compared with that of the Sanskrit Ṛg-Veda, where the corpus is also metrically diverse, on top of which, even within individual meters, there are points of considerable flexibility. Nevertheless, the Vedic differs from the Tamil in that only the former text is clearly partitioned into a small set of meters, as diagnosed largely by their consistent syllable counts and cadences across the corpus (Oldenberg 1888, Arnold 1905).

is not strongly modal (i.e. its kurtosis is low).[8] One might expect the distribution to be strongly modal if the text had a constant meter throughout, but with certain allowances, such as resolution, moraic substitution (LL = H), or catalexis, resulting in a bell curve of syllable count.

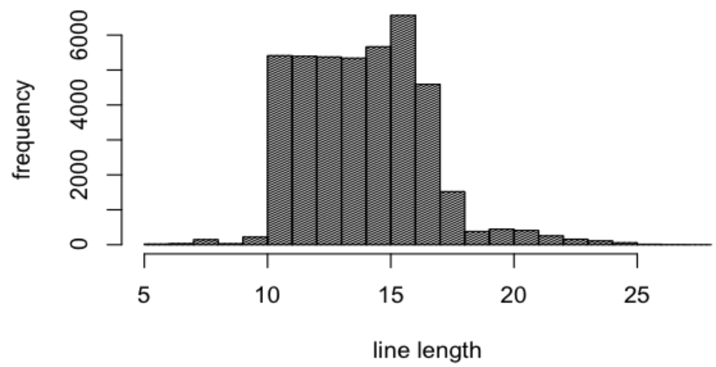

Figure 3: Histogram of line length (in syllables) in Kamban̲'s epic.

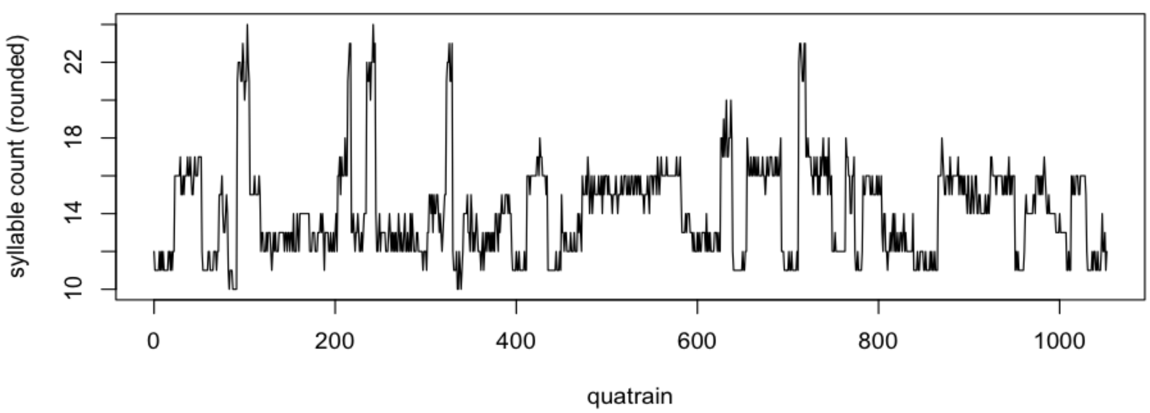

Figure 4: Line length in longitudinal perspective.

In addition to this variation in length, two lines of the same length can exhibit

[8]The conclusion is the same if one examines mora count rather than syllable count. In fact, there is generally more variation between the two lines of a couplet in mora count than there is in syllable count (standard deviation for syllables = .527, for moras = .907).

entirely different meters (assuming that they are not drawn from the same verse). Consider 16-syllable lines. The most common 16-syllable template (putting aside the final position, which is anceps, i.e., free) is HL HL HL HL HL HL HL HX ($N = 46$), that is, straight trochees (the spacing here is arbitrary, for readability). One might therefore wonder whether all 16-syllable lines are basically trochaic at some level. Indeed, the second most frequent template is also trochaic, with a single substitution in the cadence: HL HL HL HL HL HL LL HX ($N = 29$). Nevertheless, the third most frequent template bucks the trend with straight iambs: LH LH LH LH LH LH LH LX ($N = 24$). Also important here are the low frequencies of these templates: Of 42,128 lines in the text, only 6,563 (15.6%) are of modal length (16 syllables), and among modal-length lines, only 46 (0.7%) exhibit the modal (straight-trochee) template. These extremely platykurtic (low mode) distributions suggest not just metrical flexibility, but also the absence of a single meter (or few meters) underlying the text.[9]

It follows that, in describing the meter, one cannot state that the, say, seventh position of a, say, 16-syllable line is either a (preferentially) heavy or light position. In some verses (e.g. §1,119, as in figure 5), it is consistently light. In others (e.g. §2,111), it is consistently heavy. The meter of this first verse can be characterized as four periods of LLLH; the second, as two periods of LLLLLLHX. These templates are nearly rigid within their verses,[10] but rather different both from each other and from

[9]The underlying metrical disunity of the text is further reinforced by inspecting quatrains with fully rigid templates, that is, the same weight template for all four lines (ignoring word breaks). This sample is biased towards quatrains with shorter line lengths, as longer lines afford more opportunities for variation; however, for the present purposes, this bias is irrelevant. The epic contains 62 such quatrains, which exhibit no less than 32 different meters, each rigid within its quatrain. Only one template from this set, HHLLHLLHLLH, occurs more than a few times in the text.

[10]The distribution of word boundaries, though of less interest at the moment, also exhibits a tight correspondence.

the (more or less) straight iambic/trochaic patterns mentioned above. Compare also §310, in which the line regularly comprises fourteen lights followed by a single heavy.

1,119 a ilăj kulaːʋ ajilinaːn anikam eːɻ ena ʋulaːm
b n̪ilăj kulaːm makaɾa n̪iːɾ n̪eʈija maː kaʈal elaːm
c alak in maːt kaɭitʉ t̪eːɾ puɾaʋi jaːɭ ena ʋiɾaːj
d ulak elaːm n̪imiɾʋat̪eː poɾuʋum oːɾ uʋamăj jeː

LL LH LL $\boxed{\text{L}}$ H • LLL H LL LH
LL LH LL $\boxed{\text{L}}$ H • LLL H LL LH
LL H H LL $\boxed{\text{L}}$ H • LLL H LL LH
LL LH LH $\boxed{\text{L}}$ H • LLL H LLL H

2,111 a akal iʈa n̪eʈit̪ aːɭum amăjt̪ijăj jat̪u t̪iːɾa-p
b pukal iʈam emat̪ aːkum puɾăj jiʈăj jit̪u n̪aːɭil
c t̪akaʋ ila t̪aʋa ʋeːʈan taɻuʋinăj ʋaɾuʋaːn en
d ikal aʈu cilăj ʋiːɾa ʋilăjjaʋanoʈum entaːn

LL LL LL $\boxed{\text{H}}$ L • LLLL LL HH
LL LL LL $\boxed{\text{H}}$ H • LL LL LL HH
LL LL LL $\boxed{\text{H}}$ H • LLLL LLH H
LL LL LL $\boxed{\text{H}}$ L • LLLLLL HH

310 a paɾat̪anum iɭaʋalum oɾu noʈi pakiɾaːt̪’
b iɾat̪amum iʋuɭijum iʋaɾinʉ matăj n̪uːl
c uɾăjt̪aɾʉ poɻut̪inum oɻikilaɾ enăj jaːɭ
d ʋaɾat̪anum iɭaʋalum ena maɾuʋinaɾ eː

LLLL LL $\boxed{\text{L}}$ L • LL LL LLH
LLLL LL $\boxed{\text{L}}$ L • LLLL LL H
LLLL LL $\boxed{\text{L}}$ L • LLLL LL H
LLLL LL $\boxed{\text{L}}$ L • LL LLLL H

Figure 5: Three scanned verses, with the 7th position boxed.

Despite the absence of a unifying metrical template, it is clear how the weight of a syllable can be ascertained using such a corpus: One can simply check whether the syllable tends to correspond to a heavy or light in the corresponding position elsewhere in the verse, capitalizing on the templatic parallelism between lines even while lacking a model of the template generator. Light syllables tend to be paired with lights, and heavies with heavies, granting some variation due to the flexibility of the meters.

For example, in the verses cited so far, 17 instances of the diphthong [ăj] are found. As mentioned above, [ăj] is claimed to be light. We therefore expect the rime [ăj] to correspond to lights more often than to heavies. The following syllables correspond within couplets to the rime [ăj] in figures 2 and 5: [la], [t̪a], [năj-p], [lăj], [lăj], [ʋi], [la], [măj], [ɾăj], [jăj], [ʈăj], [ʋa], [ɻu], [ʋa], [ɾa], [ki], and [naɾ]. This set comprises nine lights, two heavies, and six instances of [ăj] itself. It therefore appears probable that [ăj] is indeed light. But this is not an adequate sample for statistical analysis. Analysis of the entire corpus, as in §2.3, reveals that [ăj] is intermediate between light and heavy, though closer to light. When the tests are extended to other syllable types, metrical weight is revealed to be scalar rather than categorical.

2.3 Weight as an interval scale: the diphthong [ăj]

As a first approximation (to be revised), we can observe how frequently the rime [ăj] corresponds to a light vs. heavy syllable in the corresponding position of the facing line of a matched-length couplet over the course of the entire epic.[11] We can further observe the frequency with which heavies and lights correspond to themselves in order to establish baselines (again putting aside [ăj] responsions; fn. 11). Data for

[11] A responsion of [ăj] itself is put aside as neither light nor heavy.

the whole Kamban̲ corpus are given in figure 6.[12] I adapt the term RESPONSION from Greek metrics (cf. e.g. Maas 1962:§28, Klein 2002, Nagy 2010) for this dimension of cross-line ('vertical') correspondence driven by local parallelism rather than an overarching meter, though the specific sense here is not a traditional one.

probe rime	heavy responsions	light responsions	% heavy
light (≠[ăj])	26,878	106,110	20.2 %
[ăj]	4,009	7,500	34.8 %
heavy	106,110	26,878	79.0 %

Figure 6: The weight of Tamil [ăj] as judged by responsion.

The diphthong [ăj] is thus significantly different from both light and heavy syllables ($p < .0001$ in both cases),[13], being intermediate between them. At the same time, however, [ăj] patterns as significantly closer to lights than to heavies (the percentage magnitude difference is over twice as great for the latter), supporting the traditional classification of [ăj] as light rather than heavy.

At this point, however, a confound needs to be addressed. The diphthong [ăj] is distributed differently in the lexicon from other syllable types. Most obviously, it is not found word-initially (§2.1), whereas 43% of heavies and 36% of lights are word-initial in Kamban̲'s epic. This distributional difference could potentially motivate the intermediate behavior of [ăj] in figure 6. Perhaps, as one logically possible confound, accented syllables are more likely than unaccented ones to occupy heavy positions.

[12]The duplication of 26,878 in figure 6 is not a coincidence. I treat responsion as bidirectional here, in that every 'response' syllable is also treated as a 'probe' syllable. This is not crucial; even if responsion is treated as unidirectional (the first line of each couplet being the probe, the second the response), the conclusion is the same. The values in the upper-right (light-light) and lower-left (heavy-heavy) cells are not expected to be identical, however, since the heavy-heavy and light-light responsion rates are independent.

[13]Significance for any contingency table in this thesis, unless noted otherwise, is given by Fisher's exact test two-tailed on four cells of count data.

Because accent is usually thought to be word-initial in Tamil (Christdas 1988, Krishnamurti 2003, Keane 2003, 2006), [ăj] is never accented. This negative correlation with accent could create the appearance that [ăj] is lighter than other heavies. Or perhaps initial syllables pattern as heavier than other syllables, irrespective of accent, for other phonetic (cf. domain-initial articulatory strengthening, e.g. Keating et al. 2003) or phonological (cf. initial-syllable privilege, e.g. Beckman 1998) reasons.

Yet another logical possibility is that initial syllables might coincide more often with heavy positions due purely to the distribution of word shapes in the line, independent of weight. For example, if postcaesural position tended to be a strong/heavy position, it would inflate the number of initial syllables in heavy positions, simply by virtue of the fact that only an initial syllable can occupy a postcaesural position. More generally, there are only so many ways that word shapes can be slotted into fixed-length lines, so metrical position and position in the word are not expected to be distributed fully independently, even if one ignores weight mapping preferences.

2.3.1 Controlling for word shape: holding word shape constant

In short, we should control the shape of the carrying word along with the position in that word. One approach sometimes employed by Greek metrists (Irigoin 1965, Devine and Stephens 1976, 1994) is to restrict attention to a single position in a single word shape in determining counts. By WORD SHAPE (or WORD CONTEXT), I refer to the heavy-light template of the carrying word with the syllable's position blanked out (e.g. the word context of the medial of *uɳɭumaː* is H_H). Because we are presently dealing with two syllables in a correspondence relation, I take into account both the word shape of the PROBE syllable and the word shape of its RESPONSE. The most frequent probe-response word shape pair in Kamba̱n's epic in which both probe and response are medial syllables is L_L (i.e. dibrach LLL if the medial is light and amphibrach LHL if the medial is heavy). The present corpus contains 3,844 L_L-to-L_L responses.

Figure 7 is a retabulation of figure 6 counting only syllable pairs in which both syllables occupy the context L_L. With position in the word now controlled, the conclusion is the same as in §2.3: [ăj] is significantly different from both light and heavy ($p < .0001$), being intermediate between them. At the same time, however, [ăj] is closer to lights than it is to heavies, the percentage difference again being approximately twice as great in the latter case.

probe rime	heavy responses	light responses	% heavy
light	412	956	30.1 %
[ăj]	80	92	46.5 %
heavy	1,604	412	79.6 %

Figure 7: The weight of Tamil [ăj] in dibrach-medial position.

2.3.2 Controlling for word shape: a mixed model approach

In §2.3.1, I controlled for possible confounds from word shape (i.e. from position in the word and skews in the distribution of word types in the corpus) by tabulating data from only a single position in a single frequent word shape. This solution has two shortcomings. First, it drastically reduces the amount of data that is brought to bear on the question, throwing out most relevant corpus information to observe only a small, albeit well-controlled, corner of the data. For less frequent, more specific, or more distributionally constrained syllable types, significant trends in the corpus as a whole could easily be lost on this kind of test. Second, it is not empirically clear from doing such a test whether the same result holds in other positions in other word shapes. Is it generally true that Tamil [ăj] is intermediately heavy (but closer to lights)? Or is this somehow merely a fact about these syllables in the context L_L?

These shortcomings can be addressed by scaling up to a statistical approach that takes all the corpus data into account while still controlling for word shape. One such approach is to enter word context as an effect (factor) in a regression model. I employ a generalized linear mixed model using the `lmer` method in the `lme4` package (Bates and Maechler 2009) for R (R Development Core Team 2009), fitting weights by maximum likelihood. In this case, the model is logistic, with the dependent variable being whether a syllable corresponds to a heavy (coded 1) or light (coded 0) syllable. The three rime categories in question (light, [ăj], and heavy) are given as fixed effects predicting this outcome. These effects are fixed in the sense that each is assigned a constant value in the model, as reported in the regression table in figure 8.

Every syllable in a matched-length couplet[14] in the corpus is coded for whether its correspondent in the facing line is heavy or light (presently excluding as data syllables whose correspondents are [ăj], whose weight is in question). Additionally, syllables are coded for word context (as in §2.3.1, e.g. _H represents the first position of a heavy-final disyllable), also referred to as WORD SHAPE here, which is employed as a random effect. Random effects are perhaps most familiar in linguistics as controls for by-subject idiosyncrasies in experimental modeling (Baayen 2008:§7). On the justification and implementation of mixed effects regression models in linguistics (and advantages over older approaches such as ANOVAs), see Baayen (2004, 2008:§7), Baayen et al. (2008), Jaeger (2008), Quené and van den Bergh (2008), and Levy (2010). As a random effect, word shape has a very large (perhaps, in theory, infinite) number of possible levels, given the productivity of morphology and the freedom of

[14] This underuses the data, since about half (50.5%) of couplets are not matched-length. Nevertheless, employing only this subset of the data makes it clear which syllables occupy corresponding positions in the meter, which is not always clear in facing lines of different lengths. I employ a string-alignment heuristic in §2.5 that circumvents this problem.

phonology (e.g. a proper name can have an arbitrary number of syllables), and the word shapes employed in any given corpus are only a sample, if a large one, of this population of possible word shapes.

Employing word shape as a random effect is a means of correcting for skews in the distribution of the rimes in meter that are reflexes of positional confounds, as discussed in §2.3.1, without sacrificing generality (more after the table on precisely how this works). In §2.3.1, these skews were controlled by observing only a single word context. A mixed model is now employed to generalize this control to all word shapes in the corpus.

rime	coefficient	standard error	z-value	p-value
intercept (i.e. [ăj])	-1.3016	.0926	-14.05	$< .00001$
light	$-.6238$	.0322	-19.37	$< .00001$
heavy	1.6730	.0322	51.91	$< .00001$

Figure 8: Logistic model predicting responsion from syllable type.

Figure 8 is a simple logistic regression table. Heavy and light are given as factors; in addition, [ăj], whose weight is in question, is represented by the intercept in this model. These three categories exhaust syllables; thus, any syllable that is not heavy or light is [ăj], i.e., the intercept. Though one might think of rime type as being a single factor with different rimes being conditions or levels of that factor, in this case each type is treated by the model as a (binary or Boolean) factor, so that the differences between types can be explicitly gauged. The reported numbers can be interpreted as follows (summarized by the equation in figure 9).

$$\boxed{\begin{array}{c}\text{Pr(heavy response} \mid \text{rime}_0\text{, probe}_0\text{, response}_0) = \\ 1 \,/\, (1 + \exp(-(\text{intercept} + \text{coef}_{\text{rime}_0} + \text{intercept}_{\text{probe}_0} + \text{intercept}_{\text{response}_0})))\end{array}}$$

Figure 9: Computing estimated weight from the logistic model.

To compute the estimated weight of a given syllable, sum the general intercept, the applicable random effects intercepts, and the applicable coefficient, if any, and plug the result x into the logistic (logit link) function $\frac{1}{1+e^{-x}}$, where e is Euler's number (≈2.7183). For example, consider a heavy in the frame L_L whose response is in the frame L_H. We sum the intercept (−1.3016), the relevant coefficient (1.6730), and the relevant random effects intercepts, which are not reported in the table (0.3070 for probe L_L and 0.1058 for response L_H). This gives 0.7842, which translates (through the logistic function) to $p = .6867$, or an approximately 69% chance that a heavy will correspond to another heavy, given the two word shapes in question.

To be clear, the fact that the intercept is lower than the light coefficient does not mean that [ăj] is lighter than light syllables: It merely indicates that the base case [ăj] responds to heavies less than 50% of the time, while the negative coefficient for 'light' means that 'light' responds to heavies even less often than that. Any negative coefficient is subtracted from the intercept; any positive coefficient added to it. The standard errors in the second column of figure 8 correspond to the estimated standard deviation of each factor. The coefficient is divided by the standard error to get the z-value in the third column, whose p-value (conservatively given as two-tailed, i.e. bidirectional, here) is given in the final column.

Figure 8 does not show either of the two random effects used in the model, one representing the word shape of the probe syllable ($N = 402$) and the other the word shape of the response syllable ($N = 407$). Intercepts of the 20 most frequent probe word shapes are depicted as a bar chart in figure 10 (as in Levy 2010). The y-axis is the likelihood, in log-odds space relative to the general intercept in figure 8, that 'X' in each word context 'responds with' a heavy syllable. The y-axis ranges from −0.38 (X in H_L corresponds to a heavy less often than baseline) to 0.47 (X in L_H corresponds to a heavy more often than baseline). Each code on the x-axis is accompanied by its frequency in the data employed (e.g. 7,180 probes are found in the frame H_L). Note that because accent is word-initial in Tamil (§2.3), controlling for word shape also controls for confounds from accent (among other things). For example, a monosyllabic word (#X#) is corrected according to general behavior of monosyllables; the medial of a dibrach/amphibrach (#LXL#) is adjusted to be on par with other dibrachs/amphibrachs; and so forth.

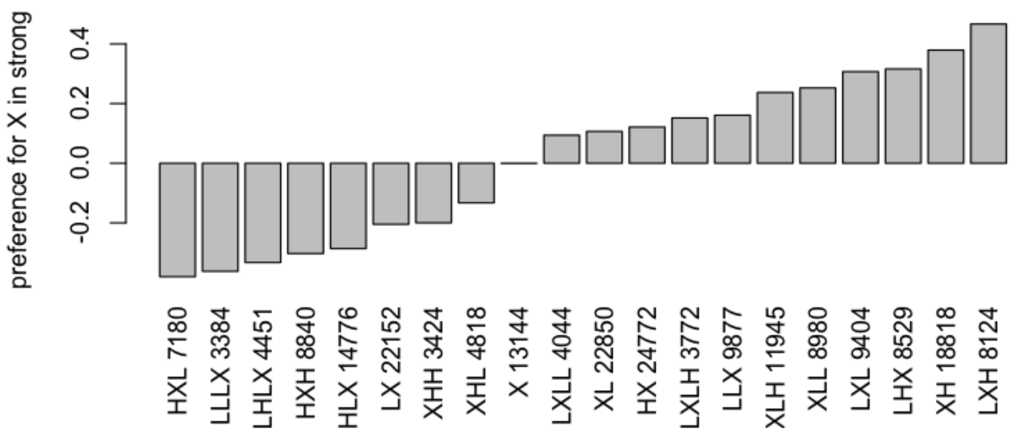

Figure 10: Intercepts of the 20 most frequent word shapes in Kambaṉ.

This logistic model reveals that, even when word shape is factored out, [ăj] patterns as intermediate between light and heavy in weight. Specifically, judging by the fixed effect coefficients in figure 8, the estimated weight (by proxy of probability of heavy responsion) of each of the three rime types is represented to scale in figure 11, in which the scale is $p \in [0, 1]$ and Δ is a difference in probability. The p-values in figure 11 are the results of the equation in figure 9 for each of the three rime types with no word shape intercepts (thus, they are disembodied p-values, in the sense that any real datum would also have to be adjusted for its word shape intercepts).

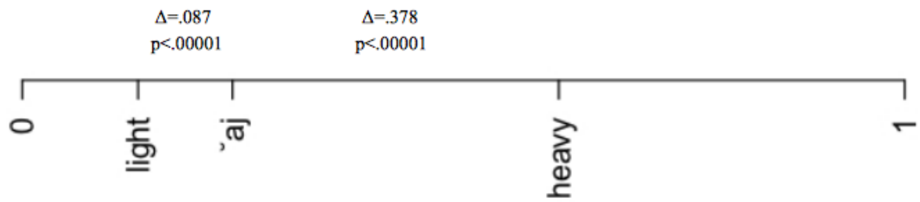

Figure 11: Estimated weights of three rime types in Kambaṉ.

Thus, with this improved methodology, the conclusion is the same as in §2.3 and §2.3.2: [ăj] is significantly different from both heavies and lights, but closer to light than heavy. Moreover, once again, the intermediate behavior of [ăj] in figure 11 is not an artifact of the distributional restriction of [ăj] to non-initial syllables. Even if all initial syllables are removed from the data, such that only non-initial lights and heavies are compared to (non-initial) [ăj], the qualitative result is the same and the numbers are only slightly different, as figure 12 reveals (regression table not shown).

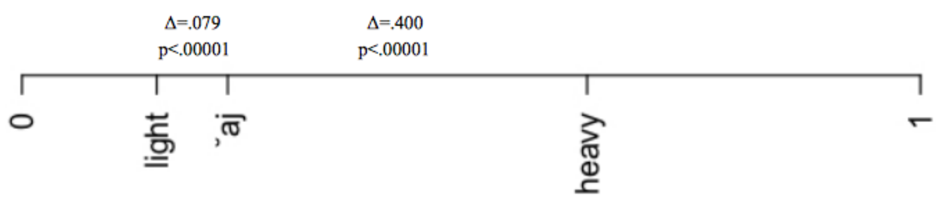

Figure 12: Estimated weights of three rime types (non-initial syllables only).

In conclusion, it appears that Tamil recognizes (at least) three weight categories. However, it is not sufficient merely to posit a three-level hierarchy for Tamil: Any adequate model of Tamil metrics must also capture the fact that the levels are not evenly spaced. In figure 11, the difference between [ăj] and heavy is roughly four times as great as that between [ăj] and light. It follows that differences are not just a matter of strict separation, but of degree. As such, they characterize an INTERVAL rather than ORDINAL scale (Stevens 1946).

2.4 A continuum of weight in Tamil

In the previous sections, I grouped Tamil syllables into three categories: light, [ăj], and heavy. I now investigate the internal structure of these categories, particularly light and heavy. As a first approximation (not controlling for word structure), figure 13 plots the weights of Tamil syllables as estimated by responsion (§2.3). The plot includes only the 115 most frequent syllables in the corpus, corresponding to a frequency cutoff of $N \geq 500$. The x-axis represents the proportion of the time that each syllable type corresponds with a heavy (where binary weight is defined as in §2.1) as opposed to light syllable. The higher the value, the heavier the syllable. The x and y values are not depicted precisely: For optimal readability, points are jittered on the y-axis and adjusted by the `pointLabel` method (part of the `maptools` package [Lewin-Koh and Bivand 2010] for R). Macrons indicate vowel length and commas alveolar (as opposed to dental) place; otherwise, the transcription is IPA.

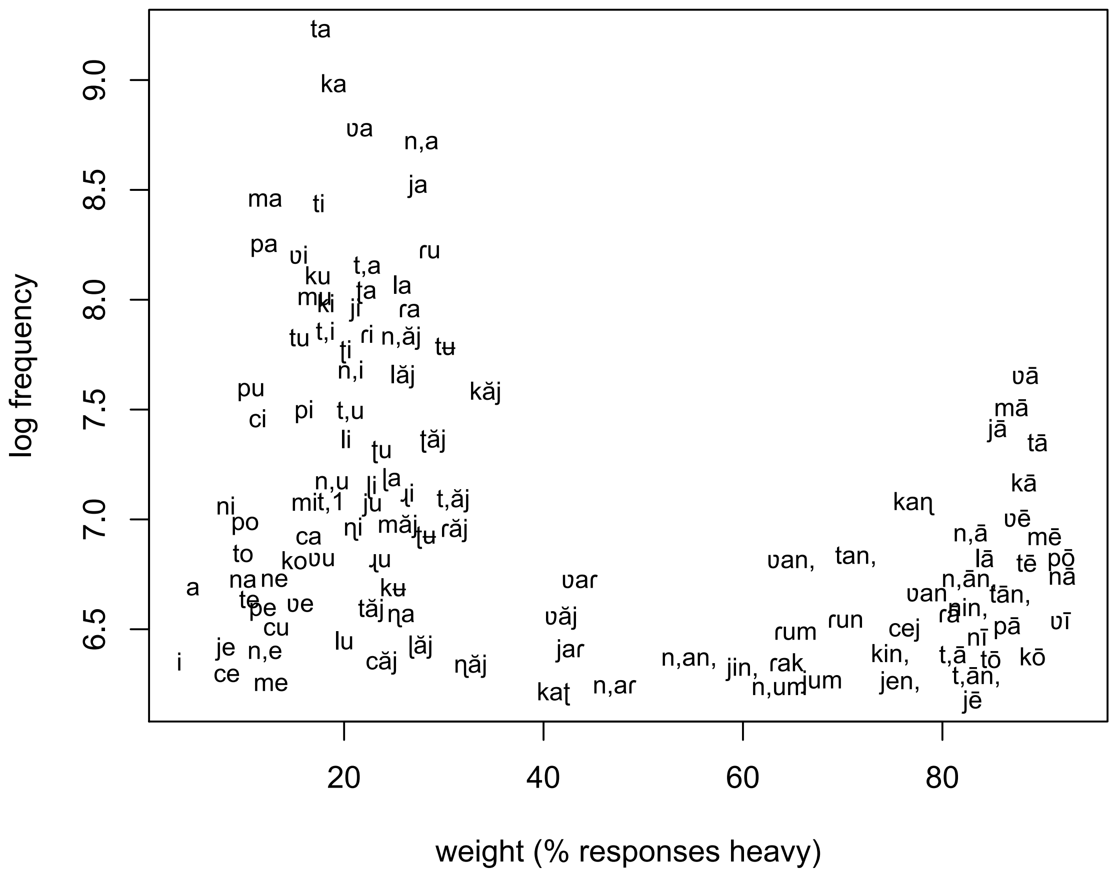

Figure 13: Responsion-estimated weights of 115 Tamil syllables.

Though one might get an initial impression of light (left) and heavy (right) clouds from figure 13, additional stratification is readily observed. Traversing the plot from left to right, syllables can be divided into at least the following relatively coherent (if overlapping) groups: $C_0\breve{V}$ (true lights), C_0ăj (light diphthong), $C_0\breve{V}R$ (R = rhotic, i.e. [ɾ] and [ɻ] in this dialect), $C_0\breve{V}N$ (N = nasal), and $C_0V{:}C_0$ (syllables with long vowels).[15] Figure 14 makes this clustering explicit by sorting figure 13 into five layers,

[15] Only two obstruent-final syllables, [kaʈ] and [ɾak], appear in figure 13. Both are on the lighter side of the heavies, but I set them aside for the moment given the small sample size.

one for each category, as labeled at left. The x-axis, which is constant through the subplots, is the same as in figure 13. The y-axes of subplots are rescaled to fit their data.[16]

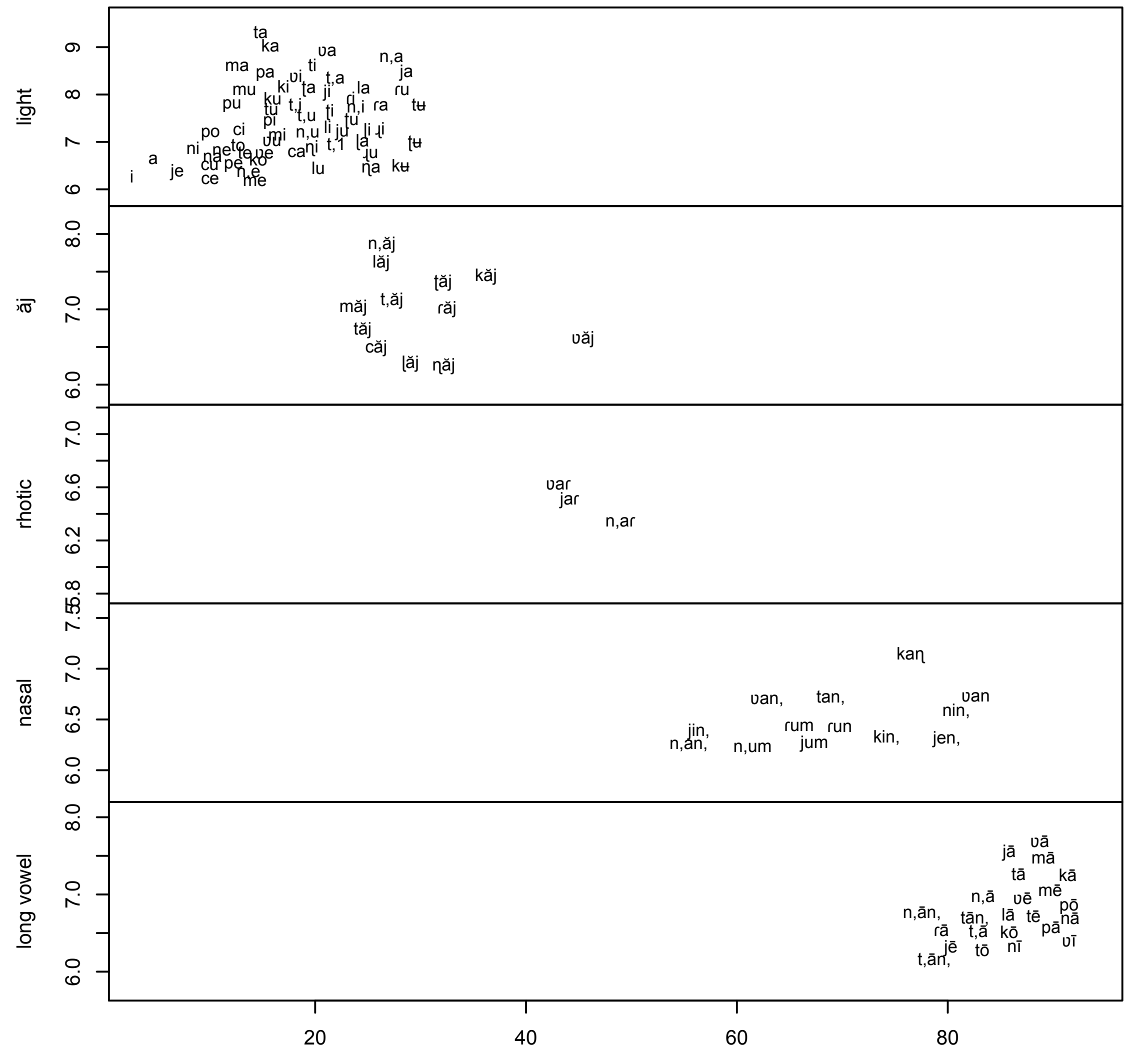

Figure 14: Figure 13 filtered into five phonological classes.

Differences between (at least) the five tiers in figure 14 are all highly significant

[16]The plotted coordinates of points are slightly discrepant between the figures due to the `pointLabel` plotting method (*op. cit.*).

($p < .00001$), as detailed along with the count data in figure 15. The columns labeled 'heavy' and 'light' stand for 'heavy responsions' and 'light responsions', respectively.

rime type	heavy	light	% heavy	p-value (vs. prev. row)
V̆	25,484	108,561	19.0 %	—
ăj	4,605	10,918	29.7 %	$< .00001$
V̆R	847	1108	43.3 %	$< .00001$
V̆N	10,793	4,480	70.7 %	$< .00001$
Vː(C)	21,099	3,345	86.3 %	$< .00001$

Figure 15: Counts and p-values for figure 14.

In conclusion, without controlling for word shape, syllable types can be stratified into at least five (perhaps many more) tiers according to their phonological characteristics. In the next section, I demonstrate that such a hierarchy persists, and can be further refined, when word shape is controlled using mixed effects regression.

2.4.1 Forward-difference coding the hierarchy in a logistic model

A mixed effects logistic regression model for syllable weight in Tamil was described in §2.3.2. I refer to this model as a LOGISTIC RESPONSION model, since it takes as a binary outcome whether each syllable corresponds to a light or heavy (0 or 1) syllable. Running this model with the five levels identified in §2.4 reveals that the hierarchy in §2.4 remains valid when (a) confounds from word shape are factored out and (b) all the data are considered, not just syllable types reaching a certain frequency threshold.

Figure 16 is the regression table. One aspect of this table must be interpreted differently from its counterpart in §2.3.2: Factors are now FORWARD-DIFFERENCE

CODED (as opposed to DUMMY CODING, the usual default in regressions).[17] Under dummy coding, each coefficient and p-value is interpreted with respect to the intercept. Thus, if a factor is significant, one can conclude only that it is significantly different from the intercept; one cannot conclude whether the factor is significantly different from any other factor. With forward-difference coding, the values of each factor are stated with respect to the previous factor in a predefined hierarchy,[18] not with respect to the general intercept. (The condition represented by the intercept itself has no comparandum other than zero.) The comparandum column makes this coding scheme explicit.

rime	comparandum	coefficient	standard error	z-value	p-value
(intercept)	[=light]	−1.1008	.0939	−10.73	$< .00001$
C_0ăj	[vs. light]	.4003	.0246	16.25	$< .00001$
$C_0\breve{V}R$	[vs. C_0ăj]	.5152	.0409	12.61	$< .00001$
$C_0\breve{V}N$	[vs. $C_0\breve{V}R$]	1.3285	.0372	35.73	$< .00001$
C_0Vː	[vs. $C_0\breve{V}N$]	.9017	.0197	45.72	$< .00001$

Figure 16: Logistic response model for five levels of weight.

The logistic equation in figure 9 is updated in figure 17 for a forward-difference coded logistic regression. By the position of the factor in the predefined hierarchy, I refer to its row in the regression table, not counting the intercept (e.g. $C_0\breve{V}R$ is the second factor in figure 16).

[17]See Venables and Ripley (2002) and *Introduction to SAS* (no author) in the references.

[18]To be clear, when I refer to a 'predefined hierarchy', I refer only to how the factors are coded, which in no way influences the findings. I intentionally choose predefined hierarchies that align with the actual numerical progression of the factors. If I had chosen to code the factors in a different order, at least one of the coefficients would be negative (i.e. lighter than the previous factor).

$$\Pr(\text{heavy response} \mid \text{rime}_0, \text{probe}_0, \text{response}_0) =$$
$$1\ /\ (1 + \exp(-(\text{intercept} + \sum_{i=1}^{n} \text{coef}_{\text{rime}_i} + \text{intercept}_{\text{probe}_0} + \text{intercept}_{\text{response}_0})))$$

where n is the position of rime_0 in the predefined hierarchy.

Figure 17: Equation for forward-difference coded logistic model.

Additionally, the R code used here for factor coding is given in figure 18, taking advantage of the `contr.sdif` method of the MASS package (Ripley 2011). Whenever forward-difference coding is employed in this dissertation, I make it explicit in the regression table by including a 'vs.' specification following each factor name, as in figure 16.

```
library(MASS); library(lme4)
categories <- c('light', 'aj', 'rhotic', 'nasal', 'long')
x$rime.f <- factor(x$rime, levels=categories)
contrasts(x$rime.f) <- contr.sdif(length(categories))
lmer(response ~ rime.f+(1|myshape)+(1|yourshape), data=x, family=binomial)
```

Figure 18: Forward-difference coding in R.

The resulting estimated weights of the five syllable classes are given to scale in figure 19. Note once again that the differences are expressed as an interval scale (with Δs being differences in probability). If the scale were expressed in terms of strict domination, significant information concerning the varying degrees of separation would be lost.

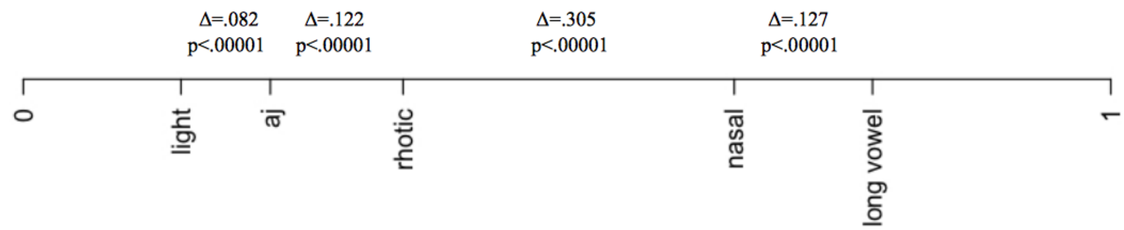

Figure 19: Relative estimated weights of five rime types in Kamban̠.

2.5 A linear propensity model of Tamil weight

Before turning to additional, more fine-grained distinctions in Tamil weight, the logistic response model in §2.4.1 can be improved upon. Potential shortcomings of this model include the following: First, it runs the risk of interference from melodic correspondence (rhyme). For one, responding syllables in line-peninitial position are usually identical, since Kamban̠ observes obligatory rhyme in that vicinity.[19] Rhyme spans often extend into additional syllables, and additional rhymes are possible elsewhere in the line. It follows that responsion between adjacent lines is not always just about weight correspondence; the poet also frequently strives to choose perceptually similar syllables, a logically possible confound for weight effects. For instance, because the diphthong [ej] is perceptually close to the diphthong [ăj], the two diphthongs might couplet-respond more often than weight alone would predict, dragging down the inferred weight of [ej] (and pulling up that of [ăj]).

Second, examining only couplet-level responsion raises issues of directionality and domain. If only the first line of each couplet is assumed to be a probe, such that responsion is unidirectional, only half the syllables in usable couplets are employed as

[19]The rhyme span begins with the first postvocalic consonant in the line and extends arbitrarily from there, with the sizes of spans ranging from the consonant alone to several syllables.

data points. If, on the other hand, responsion is treated as bidirectional, it runs the risk of oversampling the data, since each responding pair is treated as two independent data points. Furthermore, responsion is not confined to the couplet, but pervades at least the stanza (see figures 2 and 5), if not a larger group of similarly-metered stanzas. By confining responsion to the couplet, we lose information from the richer set of data bearing on the weight preference of any given position. For example, if a light syllable responds with a heavy syllable, one of the two is usually easy to discern from context as being the exception rather than the rule. For instance, the heavy in the third position in §1,119c in figure 5 appears exceptional in the context of its stanza; it is a heavy syllable in a preferentially light position. The logistic responsion model misses the fact that in a heavy-light responsion one of the two categories is typically more marked than the other.

Third, analyzing only syllables in matched-length couplets underuses the corpus data, as approximately 50% of couplets are not matched-length (fn. 14). In most cases, lines that are unmatched for length within their couplet can still be measured against other same-length lines (e.g. elsewhere in the stanza). It follows that it is not necessary to throw out data from unmatched couplets.

Finally, logistic responsion, while it controls for word shape, ignores potentially relevant information bearing on weight preferences that is tied to location in the line rather than word shape alone. For example, syllable weight is more flexible in precaesural position (usually the eighth position in a 16-syllable meter) than in anteprecaesural position, even if we confine our attention to word-final syllables. Thus, if a heavy corresponds in anteprecaesural position to lights elsewhere in the stanza, it provides more evidence that the heavy is on the lighter side than if the heavy corresponds to lights in precaesural position, which is known to be more flexible and thus not as good a diagnostic of weight preferences. Moreover, certain types of syllables might be avoided in certain parts of the line. For example, superheavy syllables are said to be avoided in cadences of Ṛg-Vedic meters, despite the presence of near-categorically heavy positions in cadences (cf. Hoenigswald 1990, 1991, 1994).

2.5.1 Estimating the weight propensity of a position

A LINEAR PROPENSITY model addresses all these potential shortcomings. The idea is that for every syllable in the corpus, one uses information from similar lines to ascertain how strongly the poet would prefer to place a heavier as opposed to lighter syllable in that context, giving a real number weight preference for that position.

One means of estimating the underlying weight propensity of a given position in a given line would be to find lines with the same length and weight template as the line in question (ignoring the position in question) and observe how often the position in question is filled by a heavy or light. In practice, however, few weight templates are repeated in more than a few times in the corpus (see §2.2). If templates are unspaced (ignoring word boundaries), 14,503 different weight templates are found, an average of 2.9 lines per template. If templates are spaced to indicate word boundaries, as in the scansions in §2.2, then 32,789 different templates are found in the corpus — almost as many templates as lines in the corpus.

The similarity criterion can be loosened up in various ways to increase the average number of comparanda.[20] The approach employed here is to require comparanda to match only with respect to a limited window surrounding the position in question, as exemplified in figure 20, in which the overall templates of the two lines, including the position in question (boxed), differ, but the window is the same, such that the

[20]Another option not pursued here (in part due to its resource intensiveness) would be to use the whole corpus as comparanda but weight each comparandum according to its similarity (e.g. Levenshtein distance) to the target line.

two lines qualify as comparanda for each other by the present criterion. Comparanda are also required to match in syllable count, so that alignment between windows is clear.[21] With a five-syllable window (two on each side of the position in question), as in figure 20, average comparanda per datum is 148.

(a)	original line (IPA):	[ilăj kulaːʋ aji li naːn anikam eːɻ ena ʋulaːm]
	spaced weight template:	LL LH LL_H LLL H LL LH
	comparandum filter:	XXXXLL_HLXXXXXXX
(b)	original line (IPA):	[aɾun̪t̪at̪i janăj jaː ɭeː jamut̪inum inijaːɭeː]
	spaced weight template:	LHLL LL_H LLLL LLHH
	comparandum filter:	XXXXLL_HLXXXXXXX

Figure 20: A five-syllable window for line comparison.

At first glance, this method might seem to have a flaw, given the discussion in §2.2: If a quatrain x exhibits the meter, say, LLLH LLL**H** LLLH LLLH and y exhibits LLLL LLL**L** LLLL LLLH, a five-syllable window renders the two meters indistinguishable in some positions, e.g. XXXXXLL_LLXXXXXX (where the blank is heavy in meter x and light in meter y). In practice, however, the loss of such contrasts due to a restricted window of analysis adds some noise to the model, but is not a confound. In the present example, the average weight of the position, if both meters were equally common, would be 50% heavy (i.e. neutral, even though the position is not, in the local context of its quatrain, actually neutral). Five-syllable windows only have an impact on the model insofar as they are consistently heavy or light. Highly variable windows, such as the (oversimplified) example presented here, are effectively washed out, in that tokens in them are assigned relatively neutral

[21] As the figure suggests, frames ignore word boundaries, but bear in mind that word context is still employed by the model as a random effect. For positions close enough to the line periphery that the window would exceed the edge, boundary symbols are used in lieu of H or L.

weight values which have little effect on their place on the inferred weight scale at the end of the analysis. Interference from rhyme is also washed out. Of the average 148 comparanda per datum, typically only one or two lines are drawn from the same quatrain as the datum, and even they have a low probability (under 10%) of being rhymes.

Given this rough estimate of the weight propensity of a position, the distribution of propensities is given as a histogram on the left side of figure 21. The histogram on the right is a sample normal distribution with the same number of data, or what one might expect to find if there were no tendency whatsoever for windows to be consistent among comparanda.[22] There is significantly more variance in the actual propensities (F-test $p < .00001$), suggesting that five-syllable windows frequently have nonrandom propensities, even though a bimodal distribution (many windows being near-uniformly light or heavy) is not observed.

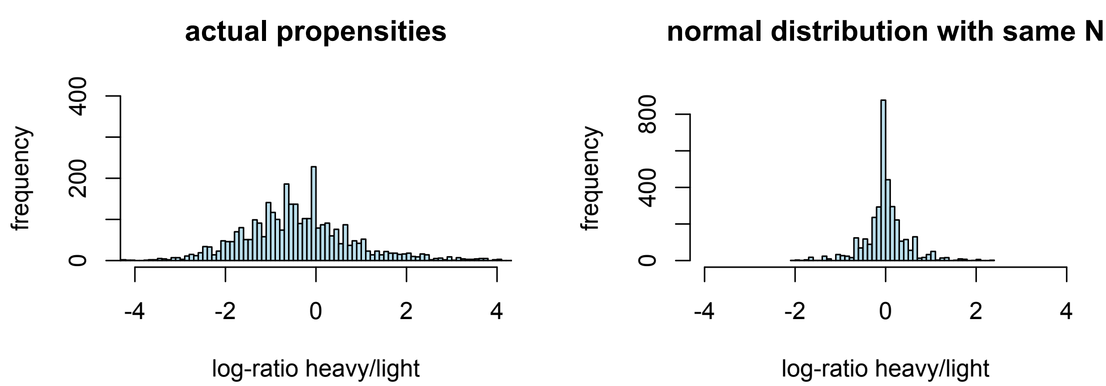

Figure 21: Histograms of observed (left) vs. normal (right) propensities.

[22] More specifically, the distribution on the right was generated by a Monte Carlo procedure, namely, running the same procedure as used to extract the distribution on the left, except replacing each actual comparandum syllable with a syllable selected at random from the corpus.

2.5.2 A ten-level linear model

In figure 16, I considered five levels of weight. Let us now double the number of factors to make ten, namely, V̆, ăj, V̆R (R = rhotic), V̆T (T = obstruent), V̆N (N = nasal), V̆L (L = lateral), V̆W (W = glide), Vː (long vowel), VːC, and VːCC. As before (§2.4.1), contrasts are forward-difference coded in the regression table in figure 22. For readability, only rime values are shown, omitting explicit indication of the optional onset (C_0). The model is still a mixed model, with a single random effect for word shape ($N = 484$).[23] Finally, the model is in this case linear rather than logistic, since the outcome (dependent variable) is a continuous real value, specifically, the log ratio of heavy to light within the set of comparanda. For example, if a datum has 112 comparanda, 100 of which contain a heavy in the relevant position, its heaviness propensity is given as $\ln\frac{100}{12} = 2.12$; it is this figure that is being estimated by the model.[24] See figure 24 for a visualization of the differences between these factors.

---

[23]Unlike the logistic responsion model above, a second random effect for response word shape is not applicable here.

[24]Given the flexibility of the meter and the number of comparanda, a zero numerator or denominator is rarely an issue in taking this log-ratio. Nevertheless, tokens with no heavy or no light comparanda were excluded for this reason, reducing the number of usable data by approximately a quarter of one percent.

rime		coefficient	standard error	t-value	p-value
(intercept)	[=V̆]	.1411	.0288	4.90	$<$ .00001
ăj	[vs. V̆]	.2896	.0074	39.26	$<$ .00001
V̆R	[vs. ăj]	.1038	.0140	7.43	$<$ .00001
V̆T	[vs. V̆R]	.5414	.0134	40.30	$<$ .00001
V̆N	[vs. V̆T]	.2381	.0070	33.92	$<$ .00001
V̆L	[vs. V̆N]	.3290	.0127	25.98	$<$ .00001
V̆W	[vs. V̆L]	.1466	.0200	7.37	$<$ .00001
Vː	[vs. V̆W]	.0128	.0168	.77	= .44
VːC	[vs. Vː]	.1130	.0090	12.58	$<$ .00001
VːCC	[vs. VːC]	.2033	.0240	8.47	$<$ .00001

Figure 22: A ten-level linear propensity model of Tamil weight.

The equations characterizing the interpretation of this regression table are given in figure 23 (cf. figure 17). The model predicts log-ratio values, as in equation (a). These can be converted to estimated probabilities, as in equation (b), which states Pr(heavy) in terms of the log-ratio obtained in (a). Most important for the present purposes, however, are the signs of the coefficients and their p-values: Any positive, significant factor can be inferred to be heavier than all preceding factors.

Let	rime_0	be a rime type (fixed effect),
	n	its position in the contrast coding hierarchy,
	probe_0	its word context (random effect), and
	comp.	its set of syllable comparanda (see text).

(a) log-ratio heavy

$$= \ln\Big(\sum_{k \in \text{comp.}} \text{if}(k{=}\text{heavy}) \;/\; \sum_{k \in \text{comp.}} \text{if}(k{\neq}\text{heavy})\Big)$$

$$= \text{intercept} + \sum_{i=1}^{n} \text{coef}_{\text{rime}_i} + \text{intercept}_{\text{probe}_0}$$

(b) Pr(heavy)

$$= \exp(\text{log-ratio heavy}) \;/\; (1 + \exp(\text{log-ratio heavy}))$$

Figure 23: Equations for forward-difference coded linear model.

The estimated weights of rimes are depicted to scale in figure 24 (where Δ, as before, is a difference in probability and the word shape intercept is left out).[25] Observe that [ăj] and V̆R are once again are closest to the lights, with a large gap up to V̆T. Short vowel plus glide (V̆W) rimes, also known as (bimoraic) diphthongs, are not significantly different from long vowels here ($p = .44$). All other contrasts are highly significant. In sum, at this point we have established nine significantly differently distributed categories of syllable weight in the Tamil corpus.

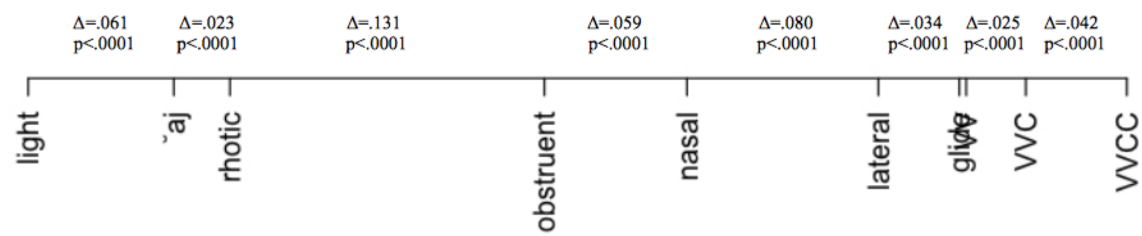

Figure 24: Relative estimated weights of ten rime types in Kamban̲.

2.5.3 The prior weight criterion is largely irrelevant

All the discussion and modeling to this point has assumed a PRIOR weight criterion for heavy vs. light syllables in Tamil, as traditionally assumed (e.g. Niklas 1988, Zvelebil 1989, Rajam 1992, Murugan 2000): $C_0\breve{V}$ is light; all other syllables are heavy. In this section, I demonstrate that assuming this criterion as a prior is not necessary to establish the hierarchies found in the preceding sections, whose qualitative features vary in only minor ways according to the choice of initial binary criterion. The only necessity is that syllables classified as light be actually lighter, *on average*, than syllables classified as heavy. In other words, the polarity of the distinction must go in the right direction, but once this condition is met, the specific cutoff between the

[25] The zero and one extrema are also now omitted in the figure. Light is aligned to $p = .270$.

heavy and light sets is largely irrelevant, and it is not necessary for all syllables to be on the right side of the divide.

For example, let us pretend that the heavy/light criterion were the following in Tamil: Syllables with short vowels are light, those with long vowels are heavy (cf. the Dravidian languages Malayalam [Mohanan 1989, Asher and Kumari 1997] and Telugu [Petrunicheva 1960, Brown 1981]). I now rerun the Tamil model from §2.5.2 with this criterion completely replacing the traditional one everywhere reference is made to binary weight (e.g. line templates, weight propensity calculation, word shapes, etc.). In other words, wherever 'H' or 'L' is employed in the model, this new criterion is now employed. The resulting hierarchy is given in the middle pair of columns in figure 25, alongside the results from §2.5.2 (with the traditional binary criterion) in the leftmost columns for comparison. As before, the coefficients are forward-difference coded (so any positive significant coefficient is heavier than the previous row, regardless of its magnitude).

As an illustration of a third possible cutoff, the rightmost columns in figure 25 show results for a prior criterion of 'light = C_0[i], heavy = other'. With this criterion, the vast majority of syllables are classified as heavy, the only exceptions being syllables ending with the short vowel [i].

prior:		light = $C_0\breve{V}$		light = $C_0\breve{V}C_0$		light = C_0i	
rime contrast		coefficient	p-value	coefficient	p-value	coefficient	p-value
ăj	vs. V̆	.290	$< .0001$	.145	$< .0001$	.097	$< .0001$
V̆R	vs. ăj	.104	$< .0001$	.173	$< .0001$	.109	$< .0001$
V̆T	vs. V̆R	.541	$< .0001$	.371	$< .0001$	.045	$< .0001$
V̆N	vs. V̆T	.238	$< .0001$	.184	$< .0001$	.105	$< .0001$
V̆L	vs. V̆N	.329	$< .0001$	.386	$< .0001$	.179	$< .0001$
V̆W	vs. V̆L	.147	$< .0001$	$-.019$	$= .37$	$-.028$	$= .006$
Vː	vs. V̆W	.013	$= .44$	.204	$< .0001$	$-.001$	$= .87$
VːC	vs. Vː	.113	$< .0001$	.313	$< .0001$	.249	$< .0001$
VːCC	vs. VːC	.203	$< .0001$	$-.213$	$< .0001$	$-.115$	$< .0001$

Figure 25: Comparing results for three prior binary criteria.

The most salient differences among the three sets of coefficients concern the placements of two categories relative to their immediate neighbors, namely, V̆W and VːCC. But even these few misalignments are never more than one step out of place with the other columns (e.g. VːCC patterns as lighter than VːC with the second and third criteria, but even in those cases, it is still heavier than Vː and all other syllables). Overall, the correlations among coefficient sets are quite strong: Over the three pairwise combinations, Pearson's r ranges from .95 to .98, all p <.00001. The correlations among the three sets are plotted in figure 26.

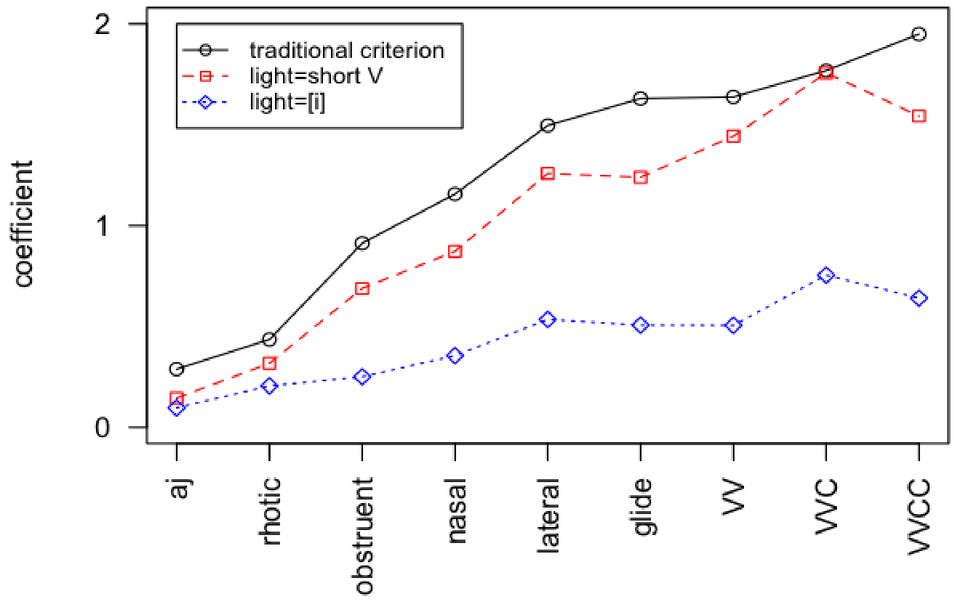

Figure 26: Scales for three priors.

Despite the general similarities of the hierarchies given different prior criteria, the differences among them, especially the handful of factor reversals reaching significance, raise the question of how authoritative we can consider any one scale, or how, for that matter, we might integrate information from multiple scales. For instance, can it really be said that V:CC patterns as heavier than V:C, when, with the two alternative priors, the reverse obtained? Likewise, V̆W was significantly lighter than V: in the 'light = $C_0\breve{V}C_0$' model. Doesn't the fact that this contrast reached significance with *any* prior suggest that it is indeed a significant difference for Kamban̠, even though this fact is only revealed by assuming a non-traditional prior? More generally, is there some way to integrate the hierarchies over possible priors? The following section addresses these questions.

2.5.4 Bootstrapping the linear model

In §2.5–2.5.3, the weight propensity of a position was estimated by taking the log-ratio of heavies to lights in the corresponding position in eligible comparanda, assuming a traditional binary weight distinction (§2.1). However, given the foregoing discussion, a more precise estimate of a position's propensity would take into account gradient weight rather than relying on one binary criterion alone. But this tack might raise concerns of circularity, since the gradient weights of syllables are precisely what is being gauged by the model. This circularity, along with the informational deficiency of assuming any binary criterion as a prior (§2.5.3), can be addressed by bootstrapping the model, that is, in this case, using the model to estimate parameters, feeding those estimates back into the model as priors, and so forth, looping until adjustments cease to significantly improve the overall likelihood of the model. This approach is a kind of Expectation Maximization algorithm (Dempster et al. 1977, Wu 1983, Hunter and Lange 2004; for a similar application in anthropology, see Holt and Benfer 2000). BOOTSTRAPPING in its most general sense refers to reusing (at least parts of) the same data sample multiple times to improve estimates based on it (Efron 1979, Efron and Tibshirani 1994, Varian 2005), including, as in the present case, resampling in order to optimize mutually-dependent variables (cf. Daland 2009).

More specifically, in this case, I begin with a particular binary criterion, running the regression model as before (§2.5.2). I then run the model a second time, using the several parameters (rime weights) estimated from the first run to more precisely determine weight propensities in the second run. Specifically, recall that weight propensity is computed by taking the average weight of comparanda (where heavy = 1, light = 0). After parameters are estimated, this average can computed from fractional weights (say, light = .25, ăj = .30, etc.). The outputs of this second model can be fed back into a third version of the model as priors, and so forth, until an acceptable degree of convergence is achieved.[26]

The change in coefficients over the course of four bootstrapped iterations is illustrated in figure 27. The plot on the left begins the first iteration with the traditional heavy/light binary prior and the one on the right begins with a rather different binary prior, namely syllables ending in [i] are light; all others are heavy (§2.5.3). After the second iteration, changes appear only slight in each plot, and the likelihood of the model ceases to improve. For example, Akaike information criteria (AICs) for the four model iterations in left plot — lower is better — are 1.76E6, 1.44E6, 1.44E6, and 1.44E6, respectively.

[26] For the sake of exposition, I update only the linear propensity calculation with the new gradient parameters, leaving line templates and word shape templates expressed in terms of binary weight.

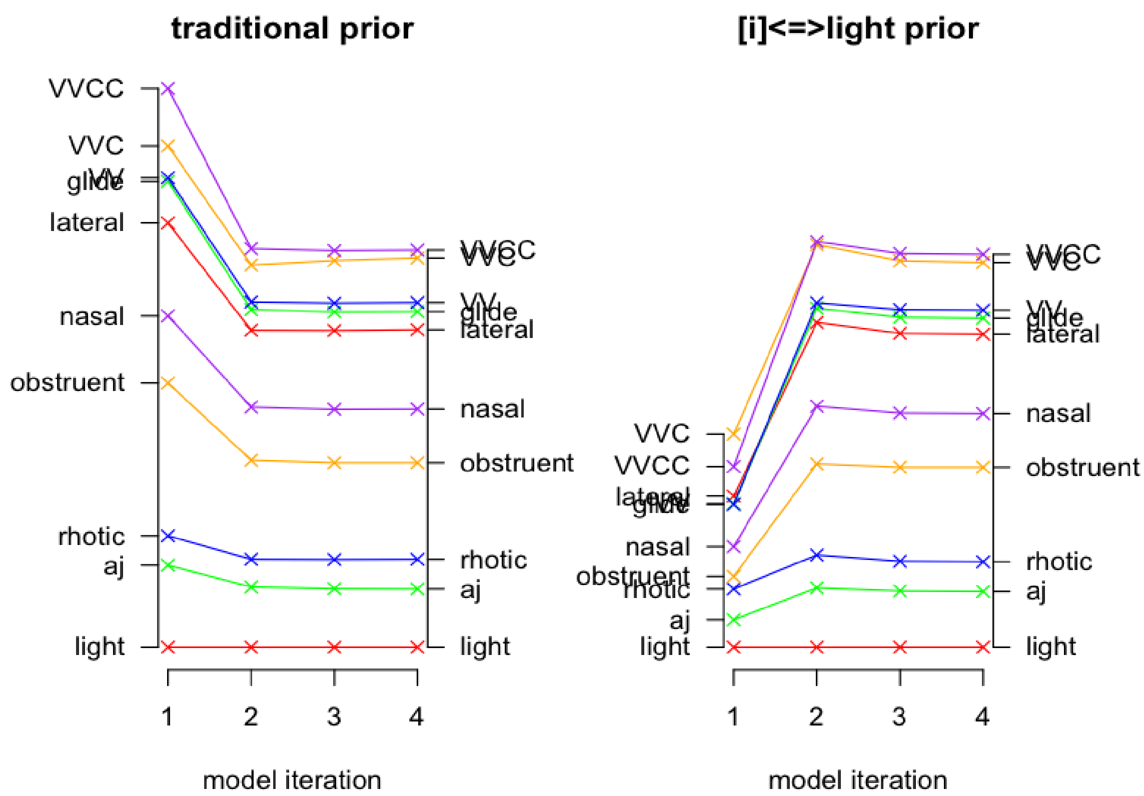

Figure 27: Evolution of rime coefficients over four iterations.

Regardless of the choice of initial prior, the final weight scale (right side of each plot) is virtually identical after a few bootstrapped iterations of parameterization. Thus, bootstrapping solves the problem of differing scales according to the prior (§2.5.3): After a few iterations, such differences wash out and the original choice of prior, even if radically different from the traditional criterion, is for all practical purposes irrelevant; the final scale can be considered more authoritative than any of the original scales. Consider, for instance, the first scales derived for each prior, on the left sides of the plots in figure 27. For a traditional prior, V:CC is significantly heavier than V:C. For the 'C_0i ⇔ light' criterion, on the other hand, V:CC is significantly lighter than V:C. However, after one or more bootstrapped iterations with either prior, the weights of V:CC and V:C converge, not being significantly different from each other (after Bonferroni correction) in either of the final scales.

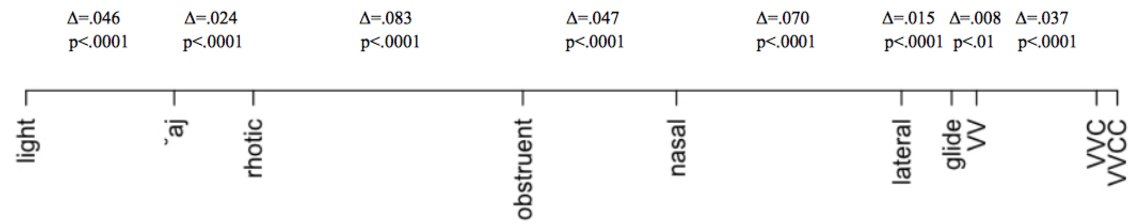

Figure 28: Estimated weights of ten rime types after bootstrapping.

The final weight scale for ten categories is given in figure 28 and the corresponding regression table is given for reference in figure 29. The contrast between Vː C and Vː CC, being nonsignificant, is not annotated in figure 28. All other contrasts are annotated and significant (V̆W $<$ Vː is borderline). As always, the scale in figure 28 is in probability space (to give an anchor, light is aligned to $p = .311$).

rime		coefficient	standard error	*t*-value	*p*-value
(intercept)	[=V̆]	−.7886	.0214	−36.89	$<$ .00001
ăj	[vs. V̆]	.1984	.0054	36.43	$<$ .00001
V̆R	[vs. ăj]	.1040	.0103	10.08	$<$ .00001
V̆T	[vs. V̆R]	.3364	.0099	33.90	$<$ .00001
V̆N	[vs. V̆T]	.1904	.0052	36.73	$<$ .00001
V̆L	[vs. V̆N]	.2805	.0093	30.04	$<$ .00001
V̆W	[vs. V̆L]	.0591	.0145	4.03	= .00006
Vː	[vs. V̆W]	.0324	.0123	2.63	= .008
VːC	[vs. Vː]	.1653	.0066	24.95	$<$ .00001
VːCC	[vs. VːC]	.0361	.0177	2.04	= .04

Figure 29: Regression table for Tamil weight after bootstrapping.

To conclude, traditional accounts of Tamil prosody posit that the poets distinguish between light (C_0V̆) and heavy (other) syllables in composing verse. I have shown that, in addition to this traditionally acknowledged criterion, the Tamil epic poet Kambaṉ is sensitive a finely articulated continuum of syllable weight. Tentatively dividing syllables into ten groups in this section to illustrate this point, nine of the groups exhibit significantly different distributions. The resulting continuum is an interval scale, in which differences are a matter not just of ordering but also of degree. This scale is revisited in §7, in which its particular features (including the perhaps surprising place of V̆R) are motivated in terms of universal principles of syllable weight; generative analysis of gradient weight mapping follows in §10. Finally, this section treats only the structure of the rime as a predictor of weight. Features of the onset are also relevant, as §12.1 argues.

I now turn to several other meters, beginning with the Latin and Ancient Greek hexameters, to demonstrate that the treatment of weight transcends binarity in them as well. The Latin and Greek differ from the Tamil in that the meters are more easily describable in terms of a uniform template for the whole text. They also exhibit categorical restrictions, in contrast to the generally more flexible character of Kambaṉ's epic.

3 The Latin and Greek hexameters

3.1 The Latin corpus and meter

I begin by describing the Classical Latin hexameter, as it is perhaps the most accessible, though chronologically it follows the Greek. As a Latin corpus, I employ Vergil's *Aeneid* (c. 25 BCE), an epic poem of 12 books composed in dactylic hexameter. The Latin text of Greenough (1900) was downloaded from the Perseus Project website (`www.perseus.tufts.edu`, accessed June 2009). This edition of the text lacks macrons (vowel length annotations), which were added manually using Pharr (1964), an edition of the first six books including the Latin text (with macrons), glosses, and translations.[27] The original text comprises 9,844 hexameter-length lines, excluding sporadic shorter lines (hexameter fragments).

The syllable weight template of the hexameter is schematized in figure 30 (Bennett 1918:245, Duckworth 1969, Allen 1973, Halle 1970, Prince 1989).[28] Each line comprises six feet (or METRA), as enumerated along the top of the figure. The first half of each metron, termed the LONGUM ('L' in the second row of the figure), is normally filled by a heavy syllable, indicated here by the en dash (–). The second half of each (non-final) metron contains either a single heavy or a pair of lights (indicated $\smile\smile$). This half is therefore sometimes called the BICEPS, labeled 'B' in the figure (West 1982, 1987). The second half of the final metron is an exception, being not a biceps but an ANCEPS ('A'), i.e., a single syllable of any weight.[29]

1		2		3		4		5		6	
L	B	L	B	L	B	L	B	L	B	L	A
$-$	$\left\{\begin{matrix}-\\ \smile\smile\end{matrix}\right\}$	$-$	$\left\{\begin{matrix}-\\ \smile\smile\end{matrix}\right\}$	$-$	$\left\{\begin{matrix}-\\ \smile\smile\end{matrix}\right\}$	$-$	$\left\{\begin{matrix}-\\ \smile\smile\end{matrix}\right\}$	$-$	$\left\{\begin{matrix}-\\ \smile\smile\end{matrix}\right\}$	$-$	$\left\{\begin{matrix}-\\ \smile\end{matrix}\right\}$

Figure 30: Hexameter template (L = longum, B = biceps, A = anceps).

The caesura (i.e. boundary between half-lines, not shown in the figure) typically

[27] After supplying diacritics by hand for this half of the text using Pharr (1964), I employed automatic heuristics to aid in extending length annotations to the remainder of the epic. Length markers were extended to words that were confidently identifiable with a unique length pattern (e.g. orthographic *non* is very frequent and always [noːn], never [non]). Phonological criteria were also employed (e.g. 'o' is almost always long word-finally, one exception being *ego* 'I'). Words in the final six books that were not reliably identified by these procedures (due to low frequency, lexical ambiguity [minimal pairs], or variability) were checked by hand. Finally, lines were retained only if they scanned properly, which provides independent confirmation of most length annotations (though vowel length in closed syllables can be neither confirmed nor disconfirmed by scansion).

[28] These illustrations ignore sporadic exceptions, such as lights in line-initial position.

[29] A light in this position is traditionally said to be *syllaba brevis in elementō longō*, or simply *brevis in longō*, i.e., read as long, which is sometimes distinguished from anceps.

falls after the third or fourth longum, or sometimes between two lights in the third biceps, but not in the middle of the line, as one might expect. Metrical ictus (downbeat) is usually claimed to fall on the longum (e.g. Bennett 1918:245, Allen 1973), though the status of the ictus is not without controversy, especially in the Greek.[30] Finally, the final biceps (i.e. that of the fifth metron) almost always (96.6% of lines in my corpus) comprises a pair of lights rather than a heavy. Some additional complications in metrification are mentioned later in this section.

Figure 31 illustrates three scanned lines from the *Aeneid*, first in orthography (with macrons) and then in IPA transcription, with closing brackets indicating the right edges of metra. See Allen (1978) for a standard account of the pronunciation of Latin. The bullet (•) marks the principal caesura. Note that underlying V#V sequences are often resolved (indicated here by replacing the first vowel with an apostrophe), as in the second metron of §1.5 and the fourth metron of §1.7.

---

[30] This position is therefore also called the THESIS ('lowering', or downbeat) as opposed to the ARSIS ('raising', or upbeat), though I avoid this terminology here, following Maas's recommendation (1962:6). These terms have been applied inconsistently in both ancient and modern times. For example, for Devine and Stephens (1975, 1994), longum is thesis and biceps is arsis; for West (1970), on the other hand, longum is arsis and biceps is thesis.

1.5 *multa quoque et bellō passus, dum conderet urbem*
múl.ta .k^{w}ó]$_{1}$k^{w}' èt .bél]$_{2}$loː .pás]$_{3}$us • dùm]$_{4}$ kón.de.re]$_{5}$t úr.bem]$_{6}$
HL L]$_{1}$ H H]$_{2}$H H]$_{3}$H • H]$_{4}$ HLL]$_{5}$ HH]$_{6}$
'he suffered many things in battle as well while he founded the city'

1.6 *īnferretque deōs Latiō — genus unde Latīnum*
iːn.fer]$_{1}$rét.k^{w}e .dé]$_{2}$oːs .lá.ti]$_{3}$oː • gé.nu]$_{4}$s ún.de .la]$_{5}$tíː.num]$_{6}$
HH]$_{1}$HL L]$_{2}$H LL]$_{3}$H • LL]$_{4}$ HL L]$_{5}$HH]$_{6}$
'and brought the gods to Latium — whence the Latin race'

1.7 *Albānīque patrēs atque altae moenia Rōmae.*
al.baː]$_{1}$níː.k^{w}e .pá]$_{2}$treː • s át]$_{3}$k^{w}' ál.taj]$_{4}$ mój.ni.a]$_{5}$ róː.maj]$_{6}$
HH]$_{1}$HL L]$_{2}$H • H]$_{3}$ HH]$_{4}$ HLL]$_{5}$ HH]$_{6}$
'and the Alban fathers and also the tall walls of Rome.'

Figure 31: Sample scansions from the *Aeneid*.

Resyllabification obtains across word boundaries in Latin, as it does in Tamil (§2.1). Consider the cadence *conderet urbem* in §1.5 in figure 31, which is scanned HLL]$_{5}$HH]$_{6}$. If the [t] were retained as the final coda of the first word rather than resyllabified as an onset, an unmetrifiable HLH sequence would arise. Basic syllabification in Latin is described as follows (e.g. Pharr 1964: appendix 1.9). V(C)V is syllabified as V.(C)V; VCCV as VC.CV, unless CC is a stop-liquid (*mūta cum liquidā*) cluster, in which case the interlude is optionally syllabified as V.CCV; finally, any more complex interlude VC$_{n}$V can be syllabified as VC.C$_{n-1}$V.

As suggested in fn. 27, lines are only retained in the final version of the present *Aeneid* corpus if they can be automatically parsed as licit hexameters (92.2% of lines could be automatically parsed according to the procedure described below). In order to facilitate metrical corpus studies, parsed lines are annotated with syllable boundaries and each syllable's metrical position. For instance, in my scheme, if I wanted to retrieve all syllables occupying the fourth longum, I could simply search for syllables tagged '41'.

Parsing is not entirely straightforward, as there are several dimensions of variability to be simultaneously countenanced. First, as already noted, each stop-liquid cluster can be scanned as either C.C (heterosyllabic) or .CC (compressed). Second, word-final syllables ending in a vowel or *m*[31] are optionally (overlooking irrelevant complications) elided preceding a vowel-initial word (bearing in mind that orthographically *h*-initial words count as vowel-initial). Elision need not be construed here as total omission, but might in some cases represent resolution (i.e. the grouping of two syllables under a single metrical position), devocalization, or other (para)phonology. Third, lines can be ostensibly hypermetric (also known as hypercatalexis or synapheia) if the line-final syllable undergoes elision with the initial vowel of the following line. Additional considerations or exceptions, such as the sporadic necessity of reading normally short vowels as long, are put aside here. The omission of lines exhibiting such additional complications only slightly reduces the size of the final parsed corpus.

The automatic parser first attempts to scan a line with the default settings (namely, no elision, no hypermetry, and stop-liquid clusters compressed). If the first attempt fails, the parser proceeds to try every permutation of the aforementioned changes until it finds a scansion. This is accomplished by identifying the site of each possible modification and assigning that site/modification a particular digit in a binary number. For example, in a line with two stop-liquid clusters and one possible elision site, the parser will try every permutation of changes from 000**b** (no changes; **b** = binary), through 001**b** (clusters unmodified, elision applied), 010**b** (second cluster divided, no elision), 011**b** (second cluster divided, elision applied), etc., up to 111**b**

[31] A word-final *m* in Classical Latin is often realized as nasalization on the preceding vowel (Pharr 1964, Allen 1978).

in this case (all three changes applied). The parser takes the first parse, if any, that succeeds in fitting the hexameter template (even if another parse were possible, it is ignored). This ensures that no parse has any more changes than are necessary to scan the line. If no parse is found, the line is rejected. In this manner, a corpus of parsed hexameters is constructed with the necessary phonological emendations (e.g. elisions and alternative syllabifications) applied. The final corpus contains 9,071 parsed lines.

3.2 The Ancient Greek corpus and meter

The Ancient Greek hexameter, as employed by poets such as Homer and Hesiod, in its basic design resembles its Latin successor, as described in §3.1 (Maas 1962, Raven 1962, Halle 1970, West 1982, Prince 1989), though there are many differences of detail (e.g. in Greek, compression of stop-liquid clusters is less frequent; in Greek, caesura is more likely to fall within the third biceps, being rare after the fourth longum; etc.).

The Ancient Greek hexameter template (omitting bridges, caesurae, etc.) is illustrated as a weighted directed graph in figure 32 (cf. figure 30, which contains somewhat less information; see also figure 35 for sample Greek scansions). The set of licit hexameters is the set of left-to-right paths through this graph. The graph is weighted in the sense that node size is proportional to the likelihood of traversal. The largest nodes (including the first 'H' [= heavy]) are traversed by approximately 100% of lines; smaller nodes are traversed less often. (The illegibly tiny nodes along the top are all 'H'.)[32] The Greek hexameter in figure 32 can be compared to its Latin

[32]Throughout this thesis, directed graphs (including finite-state automata) are produced using a combination of AT&T's FSM Library software (Mohri et al. 2009) and Graphviz, an open-source graph visualization software package (Gansner and North 1999). More specifically, graphs are created by first representing every possible line template as a non-branching path from a fixed start state to a fixed final state and then minimizing the graph using the FSM Library methods `fsmcompile`, `fsmdeterminize`, `fsmminimize`, and `fsmprint`. Figures are plotted using Graphviz.

counterpart in figure 33.

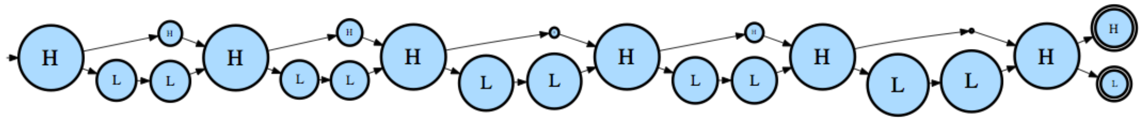

Figure 32: Weighted directed graph for the Ancient Greek hexameter.

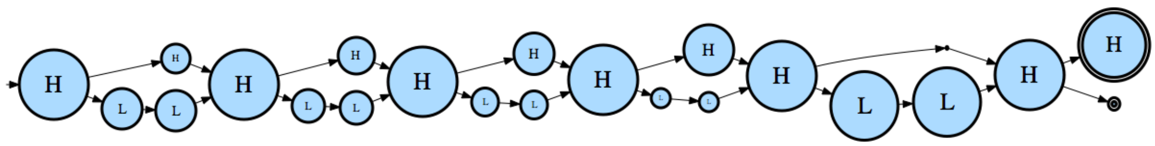

Figure 33: Weighted directed graph for the Latin hexameter.

The most salient difference between these two graphs is that metra are more likely to be spondees (– –) in Latin than they are in Ancient Greek. Figure 34 depicts the percentage dactylic (– ˘˘, as opposed to – –) across metra (skipping the third metron because it typically coincides with the caesura, and skipping the sixth metron because its second half is an anceps, not a biceps). The solid line is Greek and the dotted line is Latin. Notice that the percentage dactylic rises more or less steadily in the Greek until it reaches its maximum in the fifth metron. In the Latin, on the other hand, the proportion dactylic scoops down before spiking in the final metron.

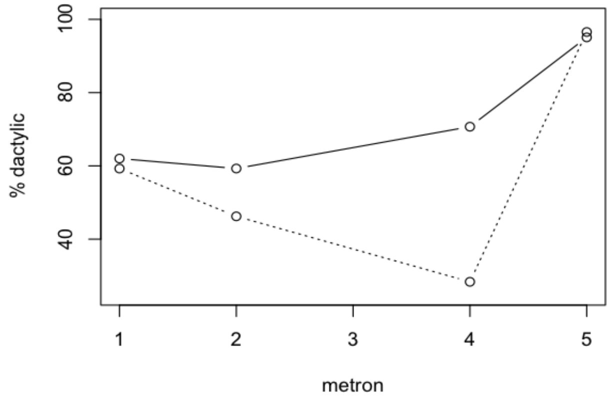

Figure 34: Percentage dactylic across metra in Greek (solid) and Latin (dotted).

My Greek corpus comprises 24,677 parsed hexameters extracted from Homer's epics, the *Iliad* and the *Odyssey* (c. 750 BCE), as available online at the Thesaurus Linguae Graecae (`www.tlg.uci.edu`, accessed April 2009). The Greek script was first transliterated into a lossless ASCII romanization. In Ancient Greek, as in Latin, vowel length is often orthographically ambiguous; e.g., α might represent either [a] or [aː], and likewise for ι and υ. (Other Ancient Greek vowels can only be read as long or short, making its script more phonetically transparent than Latin's.)

Thus, ambiguous vowels needed to be annotated for length, as they were in §3.1. First, all orthographic words containing one or more orthographically ambiguous vowels were sorted by descending frequency. The first several hundred words in this list were then hand-annotated for length. If a word could exhibit variant length patterns (readings) depending on its context/meaning, it was flagged as ambiguous. Unambiguous length patterns were then transferred into the text, and points of remaining ambiguity were examined by hand (with the help of a metrical parser). For example, the program would ask the user to check a token of Θέτι 'Thetis (personal name)', which might be [tʰétiː] (feminine dative singular, as in *Iliad* 18.407) or [tʰéti] (feminine vocative singular, as in *Iliad* 24.104). If Θέτι precedes a cluster, the program does not suggest a change to the user (since the ultima then scans as heavy regardless of whether its vowel is long or short), and the user supplies a length (or chooses to exclude the line from the corpus because the correct length is not immediately obvious). However, if Θέτι precedes a vowel- or CV-initial word, the program suggests a vowel length for the final vowel that renders the scansion correct. The user then verifies or overturns the program's suggestion.[33]

As in §3.1, following length annotation, a metrical parser was employed to automatically annotate the text for the boundaries and metrical positions of syllables. The parser is similar to that of §3.1, except without the allowances for elision or hypermetry. The treatment of stop-liquid clusters is also different in Greek. For the Homeric corpus, such clusters are treated as heterosyllabic by default (rather than compressed as in Latin). Furthermore, the possibility of compression (also known as *correptiō attica*) extends to voiceless stop-nasal clusters in addition to stop-liquid ones (Steriade 1982:186–208). Syllabification is otherwise generally comparable to that of Latin (cf. Devine and Stephens 1994:42–3, Probert 2010:100–2, Holtsmark 2010). The parser was able to scan 24,677 of the 27,758 original orthographic lines of Homer's epics. Though further refinements could be introduced to scan additional lines, the parsed corpus is already sufficiently large for the present purposes and would not benefit much from the minor additions afforded by a more sophisticated algorithm.

---

[33] I am grateful to Dieter Gunkel for extensive discussion and collaboration in the preparation of this corpus.

3.3 A weight discrepancy between longum and biceps

The hexameter template in figure 30 implies that the meter only regulates the distribution of heavy vs. light syllables, i.e., binary weight. However, as has long been acknowledged (at least among metrists working on Greek), the set of heavies occupying longa (the first halves of metra) is statistically different from the set of heavies occupying bicipitia (the second halves of metra). In particular, the biceps is sometimes claimed to be a longer position than the longum, such that the poets prefer to fill it with (on average) heavier phonological material (e.g. West 1970:186, West 1982:39, West 1987:7,22; cf. Maas 1962:§51, Irigoin 1965, Allen 1973, Devine and Stephens 1976, 1977, 1994, McLennan 1978, Aujac and Lebel 1981). West makes this claim perhaps most explicit, stating in his textbooks that '[t]he biceps, being of greater duration [than the longum –KR], requires more stuffing' (1982:39, 1987:22), and pinning the longum-to-biceps ratio at approximately 5:6 (1987:7).

West's evidence for this discrepancy includes the overrepresentation of the following four types of heavy syllables in longa relative to bicipitia (1970:186ff *et seq.*):[34] (1) $C_0V\!:$ in which V: is the result of lengthening a usually short vowel, (2) $C_0V\!:$ in which V: stands in hiatus (i.e. immediately precedes another vowel), (3) $C_0\breve{V}C.CV$ in which the cluster is a stop-liquid sequence not undergoing optional compression,[35] and (4) $C_0V\!:$ in which V: is long due to following digamma (*op. cit.*). These four types of heavies have in common that they are all on the lighter side of heavies. Some of them are so light that they could actually scan as light (if deployed in an appropriate context). For example, a short vowel followed by a *mūta cum liquidā* sequence could

[34] In consonant-vowel skeletons, C_0 indicates zero or more consonants, $\breve{V}$ a short vowel, V: a long vowel, and unadorned V a vowel of any length.

[35] See the discussion of *mūta cum liquidā* syllabification in §3.1 and §3.2.

scan as either heavy (in which case the syllabification is assumed to be $C_0\breve{V}C.CV$) or light (assuming the syllabification $C_0\breve{V}.CCV$). As another example, long vowels in prevocalic position usually undergo shortening (correption), such that they can scan as either heavy (uncorrepted) or light (correpted), as the meter requires.

In conclusion, the overrepresentation of these lighter types in the longum suggests that the longum is lighter as a metrical position than the biceps. Indeed, even in ancient times, Dionysius of Halicarnassus (*De Compositione Verborum*, §17, 1st century BCE; Allen 1973:255, Aujac and Lebel 1981, West 1982:18) cited contemporary metrical theorists (the so-called rhythmicians) as holding that the biceps was of a longer duration than the longum.

Consider $C_0\breve{V}C$ vs. C_0VV (where VV covers both long vowels, i.e. Vː, and diphthongs, which also scan as long). Both types are categorically heavy, such that either may fill either a longum or a biceps. For example, figure 35 contains two lines from the *Iliad*. The first contains a $\breve{V}C$ rime (boxed) in the second biceps; the second a VV rime in the second biceps. (Both syllables also occupy the same position in the word, each being the medial of a molossus, i.e., heavy-heavy-heavy word.)

1.128 τριπλῇ τετρ$\boxed{\text{απ}}$λῇ τ’ ἀποτείσομεν, αἴ κέ ποθι Ζεὺς
trip.lɛ̂ːj]$_1$ tet.r$\boxed{\text{ap}}$]$_2$lɛ̂ːj • t’ a.po]$_3$téj.so.me]$_4$n áj .ké .po]$_5$t^{h}i z.dèws]$_6$
HH]$_1$ H$\boxed{\text{H}}$]$_2$H • LL]$_3$HLL]$_4$ H L L]$_5$H H]$_6$

1.065 εἴ ταρ ὅ γ’ εὐχ$\boxed{\text{ω}}$λῆς ἐπιμέμφεται ἠδ’ ἑκατόμβης
éj .ta.r ^{h}ó]$_1$ g’ ew.k^{h}$\boxed{\text{ɔː}}$]$_2$lɛ̂ː • s e.pi]$_3$mém.p^{h}e.ta]$_4$j ɛː.d’ ^{h}e.ka]$_5$tóm.bɛːs]$_6$
H L L]$_1$ H$\boxed{\text{H}}$]$_2$H • LL]$_3$HLL]$_4$ H LL]$_5$HH]$_6$

Figure 35: Illustrations of biceps $\breve{V}C$ vs. VV from the *Iliad*.

VV is significantly more frequent in bicipitia than it is in longa. Specifically, the ratio of VV-to-$\breve{V}C$ is 66% greater in bicipitia than in longa in my Homeric corpus ($p < .0001$), as detailed in figure 36.

	VV rime	V̆C rime	VV:V̆C ratio
longum	75,931	58,862	1.290
biceps	19,143	8,946	2.140

Figure 36: VV:V̆C ratio in longa vs. bicipitia in Homer.

Crosslinguistically, if a language distinguishes the weight of V̆C vs. VV, it is virtually always the latter that is the heavier (e.g. Gordon 2006). These data therefore tentatively support the biceps as being a heavier position type, but certain confounds have yet to be addressed, as discussed in §3.4.

3.3.1 A note on strength vs. length in the metron

First, however, a word of clarification is perhaps in order concerning prominence vs. duration in the dactyl. As mentioned in §3.1, it is usually assumed that the longum is the strong (head) position of the metron, e.g. among generative metrists such as Halle (1970), Prince (1989), and Hanson and Kiparsky (1996). Halle (1970) goes as far as identifying the Greek dactyl as being fundamentally a trochee (a view critiqued by Devine and Stephens 1975; cf. Prince 1989). Prince's (1989) schema is given in figure 37 (F = foot [i.e. metron], S = strong, W = weak). Given this schema, one might wonder how the biceps being the weaker half of the metron is reconciled with its being the heavier half, as tentatively suggested in §3.3, and further supported in the following sections.

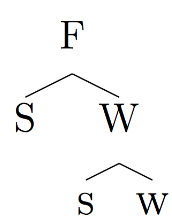

Figure 37: Prince's Greek dactyl.

For the purposes of this thesis, the weight discrepancy between the two halves of the metron is all that is relevant, while the analysis of the meter in terms of headedness and branchingness is orthogonal. On the possible dissociation of strength and length in the longum and in general (cf. English trisyllabic shortening), see especially Allen (1973:256, 286ff, 338). The typology of strophic responsion and mixed meters involving trochees or iambs being substituted for dactyls is another area in which the branchingness/headedness of feet might be probed (cf. e.g. Raven 1962:48). Third, in many of the world's meters, weak positions tend to subsume more phonological material than strong positions. For instance, in G.M. Hopkins, the syllable count of strong positions is bounded at two, but unbounded in weak positions (Kiparsky 1989). This situation might be compared to the hexameter, in which (allegedly weak) bicipitia permit two syllables, whereas (allegedly strong) longa permit only one. Fourth, poets sometimes deliberately play up such tensions. Consider, for instance, the preference for stress in the biceps in certain metra of the Latin hexameter, which Allen (1973:338) stresses should be considered counterpoint rather than conflict.

Finally, it is also possible that the assumption that dactyls are left-headed is unfounded, though I do not explore the issue here, as it is orthogonal to the present goals. For instance, West (1970) and others refer to the biceps, at least in Greek metrics, as the *thesis*, implying that ictus/downbeat falls in the second half of the metron.

3.4 Controlling for word shape

In §3.3, I showed that the ratio of VV-to-V̌C is almost twice as great in bicipitia as it is in longa, suggesting that the former are heavier as metrical positions, even though both position types are heavy in categorical terms. Nevertheless, this conclusion should not be accepted yet, given that there are possible confounds for which I have not controlled (as in §2.3.1ff). While it is possible that the hexameter template is richer than binary, as West (1970, 1982, 1987) assumes, it is also possible that the different rates with which different types of heavies fill longa vs. bicipitia is reflexive of other factors that have nothing to do with weight mapping preferences, such as position in the word or the distribution of word shapes in the line (cf. Devine and Stephens 1976, 1994).

First, position in the word is a possible confound, since VV and V̆C might tend to occur in different parts of words, and different parts of words might be treated differently by the meter, or subject to different degrees of reduction/fortition (Devine and Stephens 1994:74ff). For instance, one could imagine that word-peninitial position might tend to be more phonologically reduced than initial position (cf. initial strengthening, e.g. Keating et al. 2003). Similarly, one could imagine that penultimate position might be treated as lighter than ultimate position, given final lengthening (Lunden 2006). I am not claiming that these effects are actual confounds, merely that they are logically possible confounds which should be controlled. If VV and V̆C are distributed significantly differently with respect to position in the word, it might result in one or the other tending to pattern as lighter not because of its INTRINSIC WEIGHT as a syllable, but merely due to these sorts of positional effects.

Second, word shapes are distributed unevenly in the meter. For example, any heavy preceding a light can only occupy a longum in the hexameter (see figure 30). The initial syllable of a trochaic disyllable (i.e. heavy-light word), for one, is almost four times as likely to be V̆C as it is to be VV. This discrepancy alone might explain the higher relative rate of V̆C in longa. Even disregarding intrinsic weight, V̆C might be expected to be overrepresented in longa due to a combination of (1) the way it is distributed in words and (2) the way words are distributed in the line.

3.4.1 Holding word shape constant

One simple approach to controlling for the possible positional/distributional confounds in §3.4 is to restrict our attention to a single position in a single word shape in determining counts (cf. Devine and Stephens 1994:74). Recall from §2.3.1 that by word shape (or word context), I refer to the heavy-light template of the carrying word with the syllable's position blanked out (e.g. _H for the initial syllable of a heavy-final word). The ideal word context for such a test is both frequent and relatively distributionally unconstrained (e.g., in the present case, one that can occupy both longa and bicipitia).

Consider, for instance, the context #_H# in the Greek hexameter. HH is the third most frequent word template in the Homeric corpus, after LH and LL, which both contain lights, and are therefore more restricted distributionally. Figure 38 is a retabulation of figure 36 based exclusively on this context.

	VV rime	V̆C rime	VV:V̆C ratio
longum	6,810	3,999	1.703
biceps	3,829	1,513	2.531

Figure 38: VV:V̆C ratios in #_H# context only.

Once again, the VV:V̆C ratio is significantly greater for the biceps than for the longum ($p < .0001$), supporting the biceps being an intrinsically heavier position. However, the difference in ratios is now closer (a 49% gain for bicipitia, as opposed to the 66% gain observed in figure 36). This suggests that perhaps the initial 66% figure was in part inflated by confounds from position and word shape. However, controlling for those confounds, while attenuating the effect, does not explain it away, at least not for this context.

3.4.2 A shuffled corpus approach

In §3.4.1, I controlled for possible confounds from word shape (i.e. from position in the word and skews in the distribution of word types in the corpus) by tabulating data from only a single position in a single frequent word shape. This solution has two shortcomings, as discussed for Tamil in §2.3.2. First, it greatly reduces the quantity of data bearing on the question, throwing out most relevant corpus information (91% of syllables in §3.4.1) to observe only a small, albeit well-controlled, subset of the data. Second, it is not empirically clear from doing such a test whether the same result holds in other positions in other word shapes. Is it generally true that VV is skewed towards bicipitia, or is this somehow merely a fact about VV vs. V̆C in the context _H?

I develop two general approaches here to scale up the analysis in order to address these shortcomings: (1) a shuffled corpus (or prose) comparison method (this section) and (2) a mixed effects regression model (§3.4.3). By a SHUFFLED CORPUS, I refer to a randomly generated corpus of metrically licit nonsense. Such a corpus permits one to estimate an EXPECTED distribution of VV vs. V̆C rimes in bicipitia vs. longa, as one would expect to find if the poet had ignored the VV vs. V̆C distinction, treating weight as exclusively binary (the null hypothesis in this case). Corpus shuffling is an alternative to prose comparison models (Tarlinskaja and Teterina 1974, Tarlinskaja 1976, Biggs 1996, Hayes and Moore-Cantwell 2011) when a prose corpus is not readily available. For Ancient Greek, one might obtain prose samples from Thucydides (cf. Devine and Stephens 1976), but the text would have to be annotated for vowel length (see §3.2), which, even with help from computational heuristics, is a considerable burden. Moreover, accidental fully licit hexameters might not be very frequent in prose.

In the present case, a Perl (Wall 2010) program randomly selects words from the actual Homeric corpus and concatenates them into licit hexameters, where licit is here defined as meeting the heavy/light template schematized in figures 30 and 32.[36] Because words are selected in proportion to their frequencies, the frequency profile of word shapes is approximately identical across the real and constructed corpora. The syllabification algorithm (including the treatment of resyllabification) is also identical across the two corpora.[37] Figure 39 repeats (on the left) from figure 36 the count matrix for the actual corpus but now appends (on the right) the same matrix tabulated with respect to the shuffled corpus. Finally, the O/E (observed over expected) values in the rightmost column divide the observed VV:V̆C ratios by the expected ratios for each of the two contexts, longum and biceps.

	actual corpus			constructed corpus			O/E
	VV	V̆C	VV:V̆C	VV	V̆C	VV:V̆C	
longum	75,931	58,862	1.290	14,105	9,786	1.441	.895
biceps	19,143	8,946	2.140	5,476	3,823	1.432	1.494

Figure 39: Figure 36 adjusted for expected values.

The conclusion is the same as in the preceding sections: VV is more strongly skewed towards bicipitia than longa. The expected ratio of VV:V̆C, based on the shuffled corpus, is approximately the same in both contexts. Thus, dividing the observed ratios by the expected ratios gives a biceps/longum discrepancy that is

[36] Because a spondee (HH) is almost always avoided in favor a dactyl (HLL) in the fifth metron of the real corpus, the constructor in this case permits only a dactyl in that context.

[37] A technicality here concerns the treatment of stop-sonorant compression, which is variable in Ancient Greek. Because Homer strongly tends not to compress such clusters, the randomized corpus was constructed to exhibit a comparably low rate of compression.

approximately the same as the original discrepancy, suggesting that word shape is not a confound for the effect, at least not in the sense that taking it into account would nullify or reverse the effect.

Judging by Fisher's exact test two-tailed, VV:V̆C is significantly less than expected in the longum (O/E = .895, $p < .0001$) as well as significantly greater than expected in the biceps (O/E = 1.494, $p < .0001$). However, using a Fisher's exact test (or a chi-square test) here has two shortcomings: First, it can be used to compare observed vs. expected separately for each metrical context, but it does not compare the two contexts directly to each other. After all, significance tests can only be used to compare count matrices, not ratios. The aforementioned significance results imply, but do not technically entail, that the biceps is significantly different from the longum. Second, when using a randomized corpus as a control, one would ideally prefer more assurance that the given run is typical, since the values come out differently from one run to the next. One can imagine a situation in which the randomized corpus, however sizable, might exhibit high volatility (i.e. vary substantially from one run to the next), sometimes producing a significant result by Fisher's exact test (*vel sim.*) and sometimes not.

These shortcomings can be addressed by assessing probability using the Monte Carlo method (Metropolis and Ulam 1949, Robert and Casella 2004, Rubinstein and Kroese 2007; cf. Kessler 2001, Lin 2005, Martin 2007 for linguistic applications). In this context, probability is assessed over many iterations of random corpus construction, rather than relying on values from one iteration. In particular, if, over N rounds of corpus construction, O/E (as given in figure 39) is greater for the biceps than for the longum x times, then one can conclude that the biceps is heavier than the longum with $p < \frac{x+1}{N}$. In the present case, I test $N = 100$ constructed corpora.[38] None of these corpora exhibited a reversal of the polarity of the O/E difference in figure 39; therefore, Monte Carlo $p < .01$. The biceps continues to pattern as intrinsically heavier than the longum.

3.4.3 A mixed model approach

A second general approach to controlling for confounds from word shape is to enter word shape as a random effect in a regression model, as described in §2.3.2. In this case, the model is logistic, with the dependent variable being whether the given heavy syllable occupies a biceps (coded 1) or a longum (coded 0). The two rime categories in question (V̆C and VV) are given as fixed effects predicting this outcome. Every heavy syllable in the Homeric corpus (excluding those in line-final position, which is anceps) is coded for its rime shape (V̆C or VV) and its position (biceps or longum). Heavies with other rime shapes (e.g. VVC) are left out from the data for the moment. Additionally, every syllable in the data is coded for its position in the word and word shape (as in §3.4, e.g. _H is the first position of a spondee), which is employed by the model as a random effect (§2.3.2). The regression table is given in figure 40.

rime	coefficient	standard error	z-value	p-value
intercept (i.e. V̆C)	−14.3833	2.3965	−6.00	$< .00001$
VV	.3444	.0214	16.07	$< .00001$

Figure 40: Logistic model predicting position from syllable type.

Figure 40 does not show the random effect for word shape ($N = 114$) that is part

[38] Because corpus construction is computationally intensive and constructing 100 corpora the size of the one in figure 39 would take several hours, I use smaller constructed corpora (of 500 lines each) for the present test. This is acceptable, since making each shuffled corpus smaller only increases the likelihood of a null result, not of a false positive.

of that model. Intercepts of the 20 most frequent word shapes in the Homeric corpus are depicted as a bar chart in figure 41 (cf. figure 10 for Tamil). The y-axis here is the likelihood (in logit units) that 'X' in each word context occupies a biceps. (X is always heavy here, since only heavies are under consideration.) The y-axis ranges from -0.867 (for XL, in which X must occupy a longum) to 24.501 (for XHL, in which X must occupy a biceps). Note that the grand intercept in the regression table in figure 40 is already quite low, so an only slightly negative word shape intercept, as with the leftmost 13 intercepts in figure 41, is sufficient to predict a biceps probability of close to zero. Each code on the x-axis is accompanied by its frequency in the data employed (e.g. 12,319 heavy syllables are found in the frame _L).

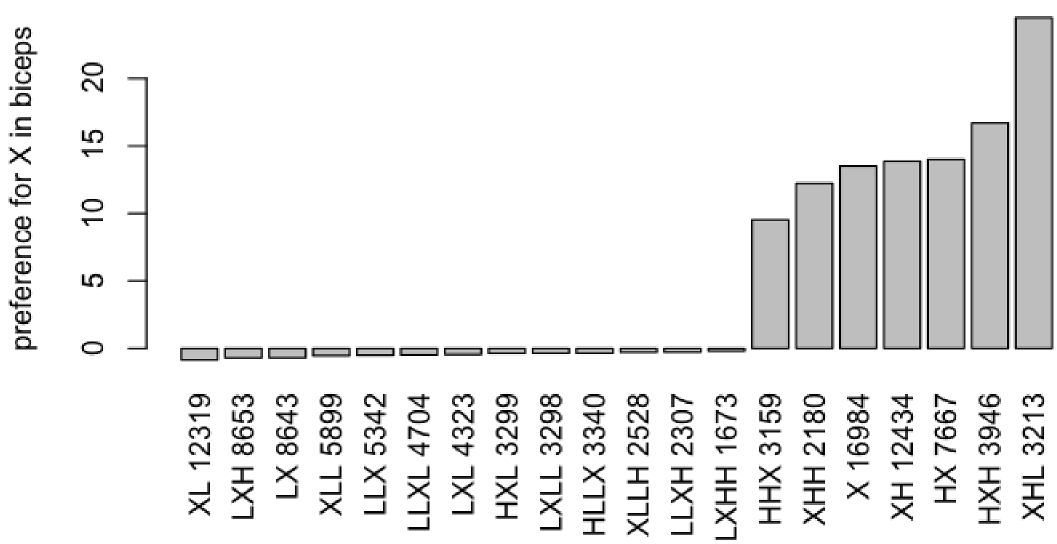

Figure 41: Intercepts of 20 word shapes (where X = heavy) in Homer.

In sum, the logistic table in figure 40 reveals that even when word shape is systematically factored out, VV exhibits a stronger skew towards bicipitia than does V̆C, supporting claims of metrists such as West (*op. cit.*) that the biceps is intrinsically heavier than the longum. Ancient Greek meter, it follows, is sensitive to not just binary weight, but to intra-heavy weight, in this case, the distinction between $\breve{V}C$ (lighter) and VV (heavier). In the following sections I demonstrate that sensitivity to intra-heavy weight in Greek and Latin is more fine-grained.

3.5 An intra-heavy hierarchy in Greek

In §3.4, I divided heavies into two types, namely, $\breve{V}C$ and VV, putting aside heavies falling into neither class. I now further subdivide the heavies, showing that Homer is sensitive to additional degrees of weight within the heavies.

Let us first split $\breve{V}C$ into two subgroups, namely, $\breve{V}$[obstruent] and $\breve{V}$[sonorant]. Based on the crosslinguistic syllable weight typology, we expect the latter to be the heavier, if any weight distinction is made between the two. For example, in Kwakwala, a $C_0\breve{V}C$ syllable is heavy iff coda C is a sonorant (Boas 1947, Zec 1995); see also Lamang (Wolff 1983, Gordon 2002). More generally, it is often observed that sonority is positively correlated with weight (e.g. Zec 1988, 1995, 2003, Morén 1999, Gordon 2006, de Lacy 2002, 2004). I also add to the model the level VVC (i.e. a closed syllable with a long vowel or diphthong). If the weight of VV is distinguished from that of VVC, we expect the latter to be the heavier, as in languages whose stress systems recognize a superheavy grade (e.g. Arabic; Allen 1983, Devine and Stephens 1994:77, Hayes 1995:67; or Pulaar; Niang 1995, Gordon 2002). More generally, more structure is expected to correlate with greater weight. In sum, we now consider a four-tiered hierarchy, stated here in order of expected weight: $\breve{V}$[obstruent], $\breve{V}$[sonorant], VV, VVC. As before, heavies not characterized by any of these four classes are put aside.

Figure 42 is a regression table for these four heavy subtypes. The model is a logistic mixed model, as in §3.4.3, in which the outcome is biceps (1) vs. longum (0) and word context is a random effect. In this figure, 'T' represents any obstruent, and 'N' any sonorant. The factors in figure 42 are forward-difference coded, as described in §2.4.1: Each coefficient and *p*-value is gauged with respect to the factor in the previous row, not with respect to the general intercept of the model.

rime		coefficient	standard error	*z*-value	*p*-value
intercept	(=V̆T)	−12.0876	1.9746	−6.12	< .00001
V̆N	[vs. V̆T]	.2726	.0342	7.97	< .00001
VV	[vs. V̆N]	.2266	.0261	8.67	< .00001
VVC	[vs. VV]	.2119	.0368	5.76	< .00001

Figure 42: Four levels of heavies in Greek (N = sonorant, T = obstruent).

The resulting estimated weights of the four syllable classes are given to scale in figure 43. In this diagram, Δ indicates a difference in probability, computed according to the coefficients in figure 42.[39] The differences in this case happen to be relatively evenly spaced from one category to the next.

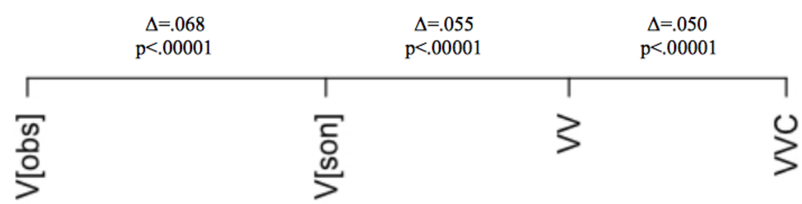

Figure 43: Relative estimated weights of four rime types in Homer.

The shuffled corpus method (§3.4.2), for its part, corroborates this four-level hierarchy. Counts of each of the types in the biceps vs. longum are given for both

[39]In calculating these baseline probabilities, I also assume a dummy word shape intercept of 12.0 in this case, so that the values are relatively centered on the [0, 1] scale, rather than compressed towards zero, as would happen if no word shape intercept were employed. Because heavies are optional in the biceps but not in the longum, all heavy subtypes are skewed towards the longum.

the actual corpus (left side) and (one iteration of) constructed corpus (right side) in figure 44. The ratios from the actual corpus are boxed in the chart, revealing a clear progression aligning with the scale reported in figure 43. After factoring out the expected values, this hierarchy remains unaltered; O/E values are given in the right-most column. The contrast between VV and VVC, however, is heavily attenuated after factoring out the expected values (Monte Carlo *p*-value undetermined).

	actual corpus			constructed corpus			O/E
rime	biceps	longum	ratio	biceps	longum	ratio	
V̆T	4,139	26,598	$\boxed{.1556}$	1,775	3,978	.4462	.3487
V̆N	4,547	21,899	$\boxed{.2076}$	1,974	4,008	.4925	.4215
VV	19,143	62,759	$\boxed{.3050}$	5,476	11,661	.4696	.6495
VVC	2,684	7,487	$\boxed{.3585}$	1,129	2,066	.5465	.6560

Figure 44: Shuffled corpus method for four levels.

3.5.1 Weight as a Hasse diagram

Figure 45 adds a fifth level to the regression analysis, namely, V̆CC. Numerically, V̆CC is intermediate between VV and VVC in weight. But it is significantly different from neither. More precisely, though its difference with respect to VVC is (slightly) less than .05 ($p = .049$), such borderline *p*-values are regarded as nonsignificant after a Bonferroni correction, according to which the α criterion (typically .05) is divided by the number of parameters (comparisons), such that the resulting criterial level for significance is .01 here. The idea behind the Bonferroni correction (or similar adjustments) is that having a greater number of factors increases the likelihood of at least one false positive. One should therefore penalize *p*-values in proportion to the number of factors.

rime		coefficient	standard error	z-value	p-value
intercept	(=V̆T)	−3.3572	1.4520	−2.31	= .021
V̆N	[vs. V̆T]	.2654	.0342	7.77	< .00001
VV	[vs. V̆N]	.2166	.0261	8.30	< .00001
V̆CC	[vs. VV]	.0054	.1009	.05	= .958
VVC	[vs. V̆CC]	.2076	.1054	1.97	= .049

Figure 45: Five levels of heavies in Greek.

Because a forward-difference coded (§3.5) regression table only permits one to assess significance with respect to adjacent factors in the predefined hierarchy, the table in figure 45 does not state whether V̆CC is significantly different from, say, V̆N. To assess the significance of V̆N vs. V̆CC, the factors need to be ordered differently, so that V̆N and V̆CC are reported as adjacent levels in the table. The ordering of factors in figure 45 also covers up the fact that VVC is significantly heavier than VV, as established in §3.5. In short, the full range of contrasts reaching significance might not be evident from a single regression table, given that p-value of each factor is gauged only with respect to the intercept (under dummy coding) or the preceding level (under forward-difference coding).

Although V̆CC is not significantly different from VV ($p = .958$), nor (after a Bonferroni correction) from V̆N ($p = .023$) it is significantly different from V̆T ($p < .0001$). This resulting weight hierarchy can be visualized as a Hasse diagram, as in figure 46. In this figure, the (implied) x-axis is weight, from lighter (left) to heavier (right). (As in figure 43, there is no y-axis.) Nodes are rimal categories. Magnitudes of difference are not drawn to scale in this case, as they were in figure 43. Any nodes that are connected by one or more solid lines are significantly different from each other (with p-values annotated on the lines).

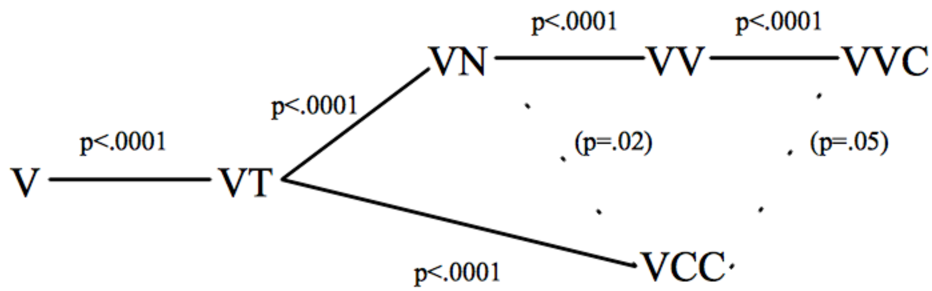

Figure 46: Five heavy types as a Hasse diagram.

Differences in figure 46 are transitive, so, for instance, V̆T being lighter than V̆N also entails that V̆T is lighter than VV. Finally, lightly dotted lines represent near-significant contrasts, as annotated by *p*-values in parentheses. V̆CC, for example, is not significantly lighter than VVC, though it comes close. A nonsignificant result should not be interpreted as positive evidence that Homer conflates the weights of V̆CC and VVC. It only signifies that the present corpus and methodology provide no evidence for such a contrast.

3.5.2 Controlling for formularity

Homer's epics are famously FORMULAIC, in that many set phrases, such as epithets, are repeated numerous times throughout the text, often filling out the same metrical positions (e.g. Sale 1993). To give one example, in my Homeric corpus, the phrase πατὴρ ἀνδρῶν τε θεῶν τε [pa.tè.r an.drɔ̂ːn .te .t^{h}e.ɔ̂ːn .te] 'father of gods and men', an epithet of Zeus, is found fifteen times in line-final position. In this section, I demonstrate that removing such formulas from the corpus does not qualitatively alter the results in §3.5.1.

Formulas could potentially be a confound in the sense that they might inflate the incidence of certain syllable types in certain positions for non-metrical reasons. For example, in the Zeus formula just cited, the rime [ɔːn] occurs twice, both times in the longum. If this formula were far more frequent than it is in reality and we were investigating the weight of, say, VVN rimes, a deceptive skew towards the longum might be found, dragging down the inferred weight of VVN. (Note that the risk of this type of confound increases as a function of both (1) the average density of formularity in the text and (2) the specificity of the categories under investigation.) On the other hand, it is also logically possible that the gradient weight hierarchy might be real but its effects confined to formulaic material, which is subject not only to compositional preferences but also to evolutionary pressures within the tradition.

A Perl program was employed to comb through the Homeric corpus and provisionally remove any (orthographic) word that was previously observed in the same metrical position(s) in the line, leaving the initial instance intact. For example, once μῆνιν [mɛ̂ːnin] is encountered line-initially, all subsequent tokens of line-initial μῆνιν are purged. This approach is deliberately overgreedy, capturing not only formulas in the usual sense, but any (orthographic) word or phrase that recurs any number of times. For instance, once τε 'and' is encountered in a given metrical position, such as the first light of the fifth biceps, all subsequent tokens of τε in that location are removed, though τε alone is not a formula.

After these changes, the number of usable heavies in the corpus is reduced by 62%. Nevertheless, even in this smaller corpus, purged of all repetitive material, the same weight hierarchy as previously reported remains in evidence, though, unsurprisingly, some of the *p*-values are attenuated (though none to the point of nonsignificance). The Hasse diagram is given in figure 47 (cf. figure 46).

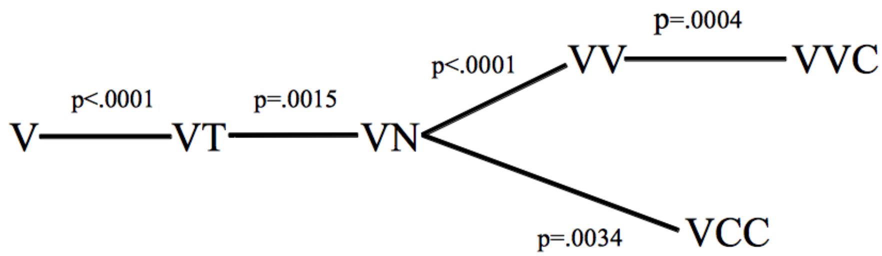

Figure 47: Hasse diagram for formula-purged Homeric subcorpus.

Since figure 47 shows results for only the nonformulaic subcorpus, we can also inspect the same results for the complement of that subcorpus, i.e., the subcorpus comprising exclusively formulaic/repetitive material (precisely the 62% of the corpus that was excluded above), as in figure 48. Once again, the same basic hierarchy emerges, though V̆CC, the least frequent of the six categories, achieves significance with respect to fewer categories in this case. Still, nothing about the Hasse diagram in figure 48 is inconsistent with that of figure 47.

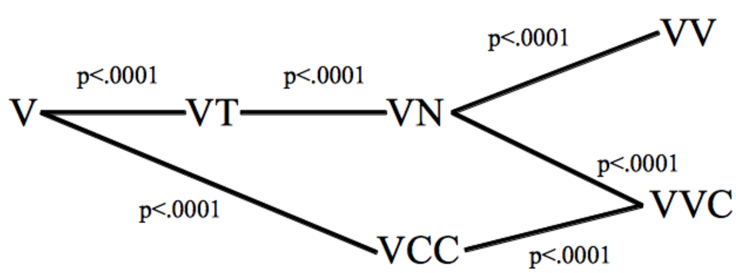

Figure 48: Hasse diagram for formulaic subcorpus.

In conclusion, the subcorpus of formulaic/repetitive material does not appear to be treated qualitatively differently from the subcorpus with formulaic/repetitive material removed. Just as the same hard metrical constraints are observed for both formulaic and nonformulaic material (e.g. the penultimate syllable of the hexameter must be heavy, regardless of the degree of formularity of the cadence), it appears that soft (preferential) metrical constraints also exert themselves in both types of material. I therefore leave in formulaic material for all the remaining corpus studies in this dissertation.

3.6 Intra-heavy weight in Latin

I argued in §3.3–3.5.2 that Homer was sensitive to a scale of intra-heavy weight, which, at this point, I have shown to be at least as articulated as V̆T < V̆N < VV < VVC (in which 'x < y' is read as 'x is lighter than y', and every contrast is significant). I now apply the same method to the Latin hexameter corpus consisting of Vergil's *Aeneid* (§3.1). A regression table is given in figure 49, set up exactly as is figure 42 for Greek (except for the stress factor; see below). As before, the binary outcome concerns whether each heavy syllable (lights are ignored) is placed in a biceps (1) or longum (0), reflecting the hypothesis that the biceps is a heavier position type, as in Greek. Word context (not shown) is included in the model as a random effect ($N = 96$), and factors are forward-difference coded (§2.4.1).

rime		coefficient	standard error	z-value	p-value
intercept	(=V̆T)	−4.5042	.7726	−5.8	< .00001
V̆N	[vs. V̆T]	.1192	.0334	3.6	= .0004
VV	[vs. V̆N]	.2249	.0297	7.6	< .00001
VVC	[vs. VV]	−.4696	.0510	−9.2	< .00001
stress		−1.7245	1.4592	−1.2	= .2

Figure 49: Four levels of heavies in Vergil (cf. figure 42 on Greek).

Unlike in the Greek model, stress level is also included as a fixed effect in the Latin model, coded as a binary factor (1 = stressed, 0 = unstressed). For uncliticized words, stress is assigned initially in all monosyllables and disyllables, while among polysyllables, it is assigned to the penult if it is heavy, and otherwise to the antepenult (Mester

1994). For cliticized words, stress is tentatively assigned immediately preceding the clitic (Jacobs 1997, Probert 2002).[40] Given that word shape (including position in the word and the word's weight template) is already controlled as a random effect, it is perhaps unsurprising that stress is nonsignificant as an additional, independent factor ($p = .2$). Note that if word shape is not included as a random effect (table not shown), the stress effect is significant in the positive direction ($z = 15.7$, $p < .0001$); that is, stressed heavies are aggregately skewed towards bicipitia.

As in Homer, the intra-heavy subhierarchy V̆T < V̆N < VV is observed (every link $p < .001$). VVC, however, is misaligned relative to the Greek (and other) results, patterning as lighter (i.e. more longum skewed) than both VV and V̆N, in the range of V̆T. A Hasse diagram for the Latin is given in figure 50. It is possible, for one, that VV undergoes (at least gradient) shortening in closed syllables, but even then VVC being lighter than V̆N is unexpected (though see below for further discussion). It is also possible, putting phonology aside, that Vergil does not distinguish between the weights of longa and bicipitia in the same way that Homer does (though in that case one would not expect any alignment between the two hierarchies). The claims cited in the preceding sections (including those of West 1970, 1982) for the biceps being heavier than the longum are based on Ancient Greek.

[40] In practice, this caveat was enforced only for *que* 'and', by far the most frequent enclitic. Other enclitics, such as *ne* and *ve*, as they are somewhat more difficult to automatically parse, are ignored here, being treated by default as part of the word; thus, the syllable preceding them is stressed iff it is heavy, with stress otherwise falling on the penult of the host.

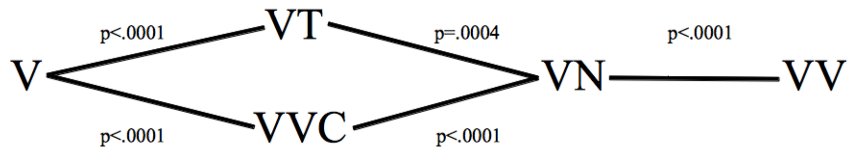

Figure 50: Hasse diagram for five rime types in Vergil.

Another consideration bearing on the weight of VVC in Latin concerns the treatment of orthographic VVN.S sequences (N = nasal, S = fricative), as in *cōnsul*. Various evidence suggests that such sequences might have been at least optionally pronounced as ṼṼ.S, i.e., as a long, nasalized vowel followed by the fricative: [kõː.sul] (Weiss 2009:61).[41] Therefore VVN.S should perhaps be classified as VV instead of VVC, as it was in figures 49 and 50. To investigate, I reran the regression in figure 49 separating VVC into three factors, namely, VVN.S, other VVN (where N is a nonnasal sonorant or a nasal coda not preceding a fricative), and VVT (as above). VVN.S and VVT were nonsignificantly different from each other and from V̆T, while VVN was nonsignificantly different from VV — in short, {V̆T, VVT, VVN.S} < V̆N < {VV, VVN}. Although it is perhaps encouraging that VVN takes its place at the top of the hierarchy once VVN.S is factored out from it, the light weights of VVT and VVN.S remain unexplained. In sum, of all the meters examined in this dissertation, Latin is the most puzzling, at least given its brief treatment here. Most of the hierarchy is consistent with the Greek and other traditions, particularly V̆T < V̆N < {VV, VVN}, but I cannot explain the relatively light treatment of the remaining subset of VVC. I leave the matter to future research.

[41]Dieter Gunkel (p.c.) alerted me to this possibility.

4 Finnish: Kalevala

4.1 Metrical and corpus preliminaries

The Kalevala is a Finnish epic poem based on Karelian folk songs compiled and edited in the nineteenth century by Elias Lönnrot (Lönnrot 1849). The meter is trochaic tetrameter, such that each line instantiates the abstract template in figure 51, in which downbeats are strong (notated ‘S’) and upbeats are weak (‘W’). Note that, as is typical, the final (eighth) position in the line, while the weak half of a metron, can be regarded as anceps, allowing a syllable of any weight.

$$\begin{array}{cc|cc|cc|cc} \multicolumn{2}{c|}{1} & \multicolumn{2}{c|}{2} & \multicolumn{2}{c|}{3} & \multicolumn{2}{c}{4} \\ \mathrm{S} & \mathrm{W} & \mathrm{S} & \mathrm{W} & \mathrm{S} & \mathrm{W} & \mathrm{S} & \mathrm{W} \end{array}$$

Figure 51: Finnish trochaic tetrameter template.

The most general rule of mapping syllable weight to strong vs. weak positions is given by Kiparsky (1968) as: ‘Stressed syllables must be long on the downbeat and short on the upbeat’ (138). (For further analysis of the Kalevala meter, see Sadeniemi 1951, Kiparsky 1968, and Leino 1986, 1994.) By ‘stressed syllables’, Kiparsky refers to the set of primary stressed syllables, which is identical to the set of word-initial syllables in Finnish. Orthographic words that comprise only one mora are treated as stressless clitics here. Short (i.e. light) syllables end in short vowels ($C_0\breve{V}$); all other syllables are long (heavy). Complex onsets are not found in Kalevala Finnish; thus, V(C)V is parsed V.(C)V, VCCV as VC.CV, and so forth. A few example lines are scanned in figure 52, in which stressed syllables and their positions are boxed. (The figure also exhibits some exceptions to the mapping rule, particularly around the beginnings of lines; see §4.1.3.)

40.221	*vaka vanha väinämöinen*							
	$\boxed{\text{S}}$	W	$\boxed{\text{S}}$	W	$\boxed{\text{S}}$	W	S	W
	$\boxed{\text{va}}$	ka	$\boxed{\text{van}}$	ha	$\boxed{\text{væi}}$	næ	møi	nen
40.224	*kalanluinen kanteloinen*							
	$\boxed{\text{S}}$	W	S	W	$\boxed{\text{S}}$	W	S	W
	$\boxed{\text{ka}}$	lan	lui	nen	$\boxed{\text{kan}}$	te	loi	nen
40.228	*ei ollut osoajata*							
	$\boxed{\text{S}}$	$\boxed{\text{W}}$	S	$\boxed{\text{W}}$	S	W	S	W
	$\boxed{\text{ei}}$	$\boxed{\text{ol}}$	lut	$\boxed{\text{o}}$	so	a	ja	ta

Figure 52: Three Kalevala lines scanned.

4.1.1 Resyllabification in the Kalevala

The status of resyllabification in Kalevala Finnish is unclear.[42] Insofar as it is only stressed syllables that are regulated for weight, as traditionally claimed (*op. cit.*), only $C_0\breve{V}C$ words are diagnostic of resyllabification. However, such words are uncommon, especially outside of the first metron. This is because, due to an independent heavy-final effect, shorter words tend to cluster at the beginning of the line (Kiparsky 1968). At the same time, it is precisely the beginning of the line that is the most flexible metrically; in fact, it has been claimed that weight mapping is completely unregulated in the first metron (Sadeniemi 1951, *et seq.*). In my Kalevala corpus (§4.1.2), outside of the first foot, $\#C_0\breve{V}C\#V$ is found only 23 times, 15 in strongs and 8 in weaks (a nonsignificant difference, and, at any rate, one that is confounded by the fact that the strong positions are on average closer to the beginning of the line than the weak positions).[43]

We can also check whether unstressed (i.e. non-initial) $\underline{\breve{V}C}\#V$, which is more

[42] I thank Arto Anttila and Paul Kiparsky (p.c.) for bringing this point to my attention.

[43] Cf. Ryan (forthcoming) on the avoidance of $\#C_0\breve{V}C\#V$ in Latin verse.

frequent, tends to be better aligned metrically with or without resyllabification. A syllable is ALIGNED if its weight agrees with its position's preference, being a heavy in a strong position or a light in a weak position. I consider here only the second half of the line (positions 4 through 7), which is more tightly regulated. Without resyllabification, $\underline{\breve{\mathrm{V}}\mathrm{C}}\#\mathrm{V}$ is aligned 446 times and non-aligned 113 times. With resyllabification, these numbers are simply reversed: $\underline{\breve{\mathrm{V}}\mathrm{C}}\#\mathrm{V}$ is aligned 113 times and non-aligned 446 times. Thus, unstressed word-final $\breve{\mathrm{V}}\mathrm{C}$ is significantly ($p < .0001$) more aligned without resyllabification than with it. Unstressed syllables have been claimed to be unregulated by the meter, making this skew perhaps surprising, but I find this claim to be unsupported (see §4.3). I therefore tentatively do not resyllabify in the following corpus studies, though in most cases this decision is irrelevant. For example, in all the studies of stressed syllables, the issue of resyllabification is virtually moot.

4.1.2 Composition of the corpus

My Kalevala corpus comprises 15,846 octosyllabic lines extracted from the text as available online at Kaapelisolmu (`www.kaapeli.fi/maailma/kalevala`, accessed June 2009). A theoretical issue bears on the construction of this corpus: Kiparsky (1968) maintains that the Kalevala is metrified according to derivationally intermediate rather than surface phonology. Specifically, he claims that five phonological rules are ordered after the metrically relevant level. I constructed the present corpus so that it is moot for the present purposes whether Kiparsky (1968, 1972) is correct in his theory of presurface metrification (cf. Manaster Ramer 1981, 1994, Devine and Stephens 1975; philological issues are also raised, cf. Lauerma 2001). I accomplished this by retaining only lines whose surface forms either match their alleged metrification forms or else depart from them only in irrelevant ways. First, two of the rules, contraction and apocope, affect syllable count, so by taking only surface octosyllables, there is no possibility that either of these rules applied. If they had, the surface form would have fewer than eight syllables. Two other rules, vowel and consonant gemination, are addressed by excluding all lines whose surface form contains a sequence that might have been an outcome of either rule. These exclusions reduce size of the corpus by 11% (from 17,890 lines to the present 15,846). Finally, the diphthongization is moot for the present purposes because diphthongs and long vowels are collapsed in the following tests.

4.1.3 The distribution of exceptions in the meter

As Kiparsky (1968) observes, and as figure 52 reinforces, the Kalevala mapping rule (S⇔H; W⇔L; figure 51) has many exceptions, but is enforced increasingly stringently towards the end of the line, almost to the point of categoricality in the 7th position (recall that the 8th position is anceps). The numbers of exceptions to the mapping rule (i.e. 'non-aligned' syllables in the parlance of §4.1.1) in each non-final position is given in figure 53, which considers only stressed (word-initial) syllables. As the plot in the figure illustrates, the decline in exceptions is almost one-to-one (i.e. slope = −1) if one counts exceptions on a logarithmic scale. This flexibility of the meter is relevant for the corpus studies below.

position	exceptions
1	5,146
2	1,165
3	127
4	126
5	73
6	27
7	9

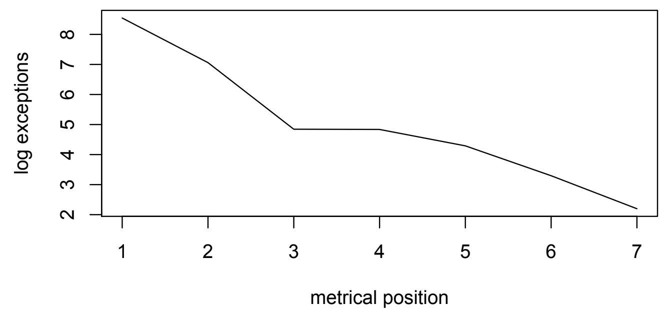

Figure 53: Distribution of exceptions to the Kalevala weight mapping rule.

4.2 Intra-heavy weight in Finnish

4.2.1 Strong vs. weak positions

Examining the exceptions to the Kalevala weight mapping rule (§4.1.3) more closely, I find that sensitivity to an intra-heavy scale of weight is affecting the poets' versification choices. In particular, poets tend to prefer lighter heavies in weak positions than they do in strong positions. Moreover, analyzing the relative skews of different syllable types between strong and weak positions permits one to derive an intra-heavy hierarchy of weight.

Let us begin with V̆C vs. VV, as we did in §3.3 with Ancient Greek. VV subsumes both long vowels and diphthongs. A contingency table for the incidence of each type in strong vs. weak positions is given in figure 54. Counts are based on all stressed syllables except those from the line-peripheral positions (1 and 8), which are (loosely speaking) anceps. Counts for light syllables are given in the first row of the table for comparison. Although both V̆C and VV rimes are found in strong positions over 90% of the time in this data set, the strong:weak ratio for VV is over twice as great as that of V̆C ($p < .00001$). Put differently, despite V̆C and VV being roughly equally common in word-initial (i.e. stressed) position (counts in figure 54), if the poet chooses to place a stressed heavy syllable in a weak position, he or she is over twice as likely to choose one with a $\breve{\mathrm{V}}$C rime over one with a VV rime (compared to the baseline ratio from strong positions).

	strong	weak	% strong	strong:weak ratio
$\breve{\mathrm{V}}$	270	9,388	2.8 %	.03
$\breve{\mathrm{V}}$C	13,438	871	93.9 %	15.43
VV	10,935	307	97.3 %	35.62

Figure 54: Overrepresentation of $\breve{\mathrm{V}}$C in weak positions.

At the same time, however, $\breve{\mathrm{V}}$C and VV are much closer to each other in weight than either is to $\breve{\mathrm{V}}$. Weight is therefore once again being treated as an interval scale (§2.3): Multiple levels are significantly differently distributed from each other, but the levels are far from evenly spaced. In the present case, the gap between $\breve{\mathrm{V}}$ and $\breve{\mathrm{V}}$C, straddling what is traditionally regarded as the heavy-light cutoff, dwarfs the gap between $\breve{\mathrm{V}}$C and VV, which are both considered categorically heavy.

As further support of the poets' treatment of $\breve{\mathrm{V}}$C heavies as lighter than other heavies, across the positions of the line, the percentage of heavies that are $\breve{\mathrm{V}}$C exhibits a negative correlation with the heaviness propensities of positions. The solid line in figure 55 represents weight propensities of positions (i.e. percentage of syllables in each position that are heavy), whose peaks are in odd (strong) positions and whose valleys are in even (weak) positions. The dotted line represents $\breve{\mathrm{V}}$C share (i.e. percentage of heavies that are $\breve{\mathrm{V}}$C), which exhibits the converse pattern — peaks in even (weak) positions and valleys in odd (strong) positions. In short, the poets prefer lighter (e.g. $\breve{\mathrm{V}}$C) heavies in weak positions.

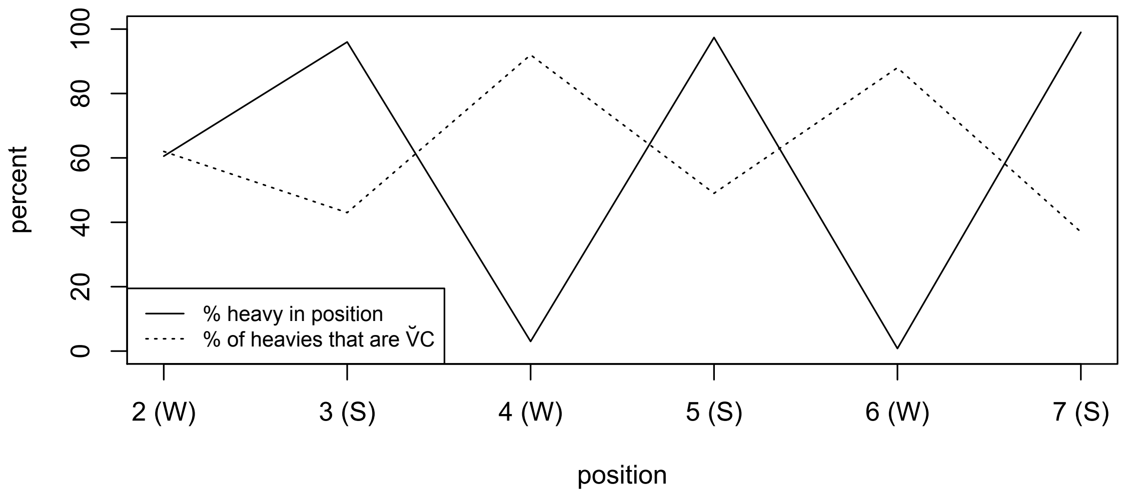

Figure 55: Negative correlation of V̆C share with positional strength.

The significantly different treatment of V̆C vs. VV, as well as a more articulated intra-heavy hierarchy, are confirmed by better controlled tests, including both a prose comparison model in §4.2.2 and a logistic model with word shape as a random effect in §4.2.3.

4.2.2 Prose model for strong vs. weak positions

To confirm that this trend is not a reflex of lexical statistics, emerging from the distribution of heavies in the lexicon even absent any active subcategorical preferences on the part of the poets, I conduct a prose comparison test (Tarlinskaja and Teterina 1974, Tarlinskaja 1976, Biggs 1996, Hayes and Moore-Cantwell 2011; cf. the shuffled corpus test for Homeric Greek in §3.4.2). My source of prose is the Finnish translation of the Bible published in 1776 (*Vuoden 1776 Raamattu*, as downloaded from `fin.scripturetext.com`, March 2010). For each Kalevala line, a Perl script selects a random octosyllabic phrase (i.e. contiguous sequence of one or more whole words) from the prose corpus whose heavy/light template and word boundary locations match those of the Kalevala line. In this manner, a 'fake Kalevala' is constructed in which categorical weight and word boundaries are distributed exactly as they are in the real Kalevala.

Figure 56 divides heavy rimes exhaustively into four types, namely, $\breve{\mathrm{V}}\mathrm{C}$, VV, $\breve{\mathrm{V}}\mathrm{C}_2$, and VVC_1. Each rime type anchors three bars. The leftmost (dark gray) bar is the percentage of the time that the rime occupies strong positions in the actual Kalevala (considering only positions 2 through 7, as in §4.2). The second (light gray) bar represents an expected percentage strong, based on the Kalevala-like corpus of octosyllabic prose extracts described in the previous paragraph. Finally, the rightmost (black) bar is the percentage difference between observed and expected. The numbers are given below the bar chart in figure 57. All percentages are based solely on stressed syllables in both corpora; unstressed syllables are ignored here.[44]

[44]The bars for both the real and fake corpora are all above 85% in figure 56. Bear in mind that, as mentioned, the fake corpus was constructed to match the real corpus exactly in the distribution of categorical weight and word boundaries. It is therefore guaranteed that stressed (i.e. word-initial) heavies of all types will be overwhelmingly skewed towards downbeats in the fake corpus, as that is the situation in the real corpus. All that is of interest in figure 56 is the relative skews between different heavy subtypes, since that was left uncontrolled in the construction of the fake corpus, allowing the prose tendencies to assert themselves.

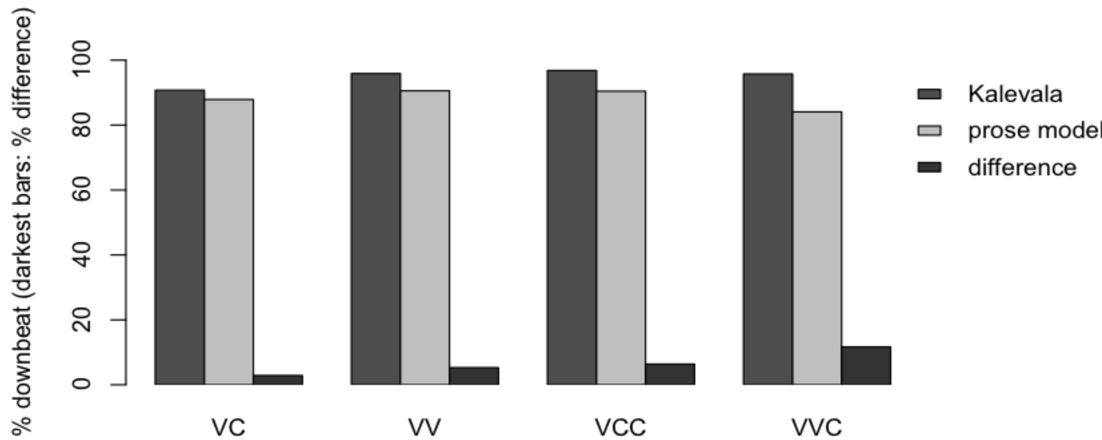

Figure 56: Observed vs. expected alignments for Finnish rimes.

	% strong observed	% strong expected	observed−expected
$\breve{V}C$	90.76 %	87.90 %	2.86 %
VV	95.87 %	90.57 %	5.30 %
$\breve{V}C_2$	96.83 %	90.44 %	6.39 %
VVC_1	95.76 %	84.07 %	11.69 %

Figure 57: Data corresponding to figure 56.

Comparing the relative observed and expected values here highlights the importance of controlling for word shape (see also §2.3.1 and §3.4). Observed values on their own terms can be misleading. For example, VVC_1 in figure 57 is slightly less skewed towards strong positions than $\breve{V}C_2$. However, after correcting for expectation, the reverse obtains: VVC_1 patterns as heavier than $\breve{V}C_2$, as one might expect given the greater sonority of the former (Gordon 2006). Moreover, the expected values exhibit differences among themselves that cannot be written off either to chance (given their p-values, e.g. $p < .00001$ for $\breve{V}C$ vs. VV) or to metrical constraints (since the corpus is extracted from prose). Rather, these differences follow from word shape confounds, such as unevenness in the distribution of rimes in the lexicon. For example, irrespective of the meter, $\breve{V}C$ share in word-initial heavies, the primary seat of vowel length contrasts in Finnish, is roughly half that of peninitial heavies (e.g. judging by token counts in the Kalevala, only 46% of initial heavies are $\breve{\text{V}}$C, compared to 84% of peninitial heavies). It follows that the $\breve{\text{V}}$C share in, say, the second metrical position is predicted to be greater than that of, say, the first simply by virtue of this skew in the lexicon. After all, the second position comprises both word-initial and penitial syllables, whereas the first position contains only word-initial syllables. A prose or shuffled corpus comparison is one means of controlling for such confounds.

4.2.3 Logistic model for strong vs. weak positions

Figure 58 is a logistic regression table for intra-heavy weight in Finnish, considering only differences of skeletal structure, specifically, the four heavy rime types discussed in §4.2.2: $\breve{\text{V}}$C, VV, $\breve{\text{V}}\text{C}_2$, and VVC_1. The binary outcome in this case concerns whether each datum is placed in a strong (1) or weak (0) position. Thus, given that strong positions attract weight, positive coefficients represent added weight. As before, only line-medial stressed syllables are taken as data. Word shape is entered in the model as a random effect ($N = 55$), as in §2.3.2 and §3.4.3. Finally, as in §2.4.1, factors are forward-difference coded, to be interpreted in row-wise succession rather than with respect to the intercept.

rime		coefficient	standard error	z-value	p-value
intercept	(=$\breve{\text{V}}$)	1.958	.521	3.76	= .0002
$\breve{\text{V}}$C	[vs. $\breve{\text{V}}$]	6.689	.086	77.40	< .00001
VV	[vs. $\breve{\text{V}}$C]	.870	.076	11.42	< .00001
VC_2	[vs. VV]	.186	.323	.58	= .57
VVC_1	[vs. VC_2]	.649	.333	1.95	= .05

Figure 58: Logistic model for skeletal rime structure in Finnish.

The following hierarchy is highly significant (each link $p < .00001$): $\breve{\text{V}} < \breve{\text{V}}\text{C} <$

$VV < VVC_1$ (this last pair is not explicit in figure 58 but can be inferred from it, given the coefficients and their standard errors, or checked explicitly by leaving out $\breve{V}C_2$ as a factor). $\breve{V}C_2$, for its part, is not significantly different from VV; however, it is significantly heavier than $\breve{V}C$ ($p = .002$) and near-significantly lighter than VVC_1 (two-tailed $p = .05$). Note that $\breve{V}C_2$ is relatively rare, comprising 1.7% of the heavies. The findings are summarized as a Hasse diagram (§3.5.1) in figure 59 (edge lengths not to scale).

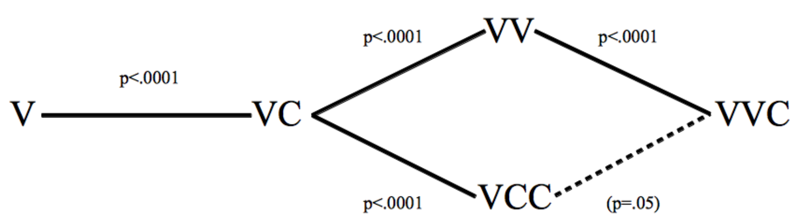

Figure 59: Hasse diagram for Finnish rime skeletons.

If VV is split into surface long vowels and surface diphthongs (though cf. §4.1.2 for a caveat concerning diphthongs), both are independently significantly heavier than $\breve{V}C$ and significantly lighter than VVC_1, as figure 60 illustrates. The unlabeled solid edges are all $p < .00001$. They are not, however, significantly different from each other (Vː being slightly heavier than the diphthongs numerically). As before, only data from primary-stressed syllables are considered here.

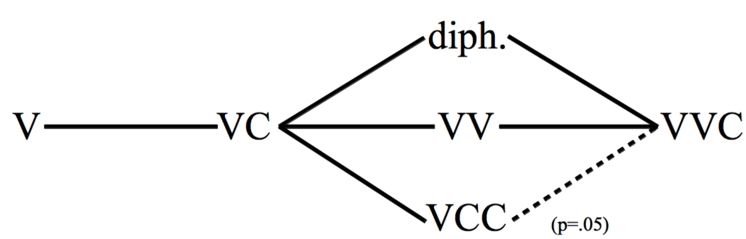

Figure 60: Hasse diagram for Finnish rimes (bifurcating VV).

4.3 Intra-heavy weight in unstressed syllables

The usually cited rule for Kalevala weight mapping (as in, e.g., Kiparsky 1968 and Hanson and Kiparsky 1996; see §4.1) applies only to primary-stressed (i.e. word-initial) syllables, suggesting that the distribution of weight is metrically unregulated in non-word-initial syllables. I find this assumption to be false, as I demonstrate here first with prose comparison (§4.3.1) and then with logistic modeling (§4.3.2).

4.3.1 Testing unstressed syllable alignment against prose

First, as in §4.2.2, an artificial Kalevala is constructed by taking each line of the actual corpus and randomly selecting an octosyllabic phrase from prose matching it on selected criteria (i.e. controls). In this case, I control for boundary location and the weights of all stressed syllables, leaving the weights of unstressed syllables uncontrolled. For example, a Kalevala line of the form LL#HLHL#HL (H = heavy, L = light) is matched by any phrase of the form LX#HXXX#HX. The distribution of unstressed heavies and lights in this prose-based Kalevala can then be compared to that of the actual Kalevala as a means of testing whether unstressed syllables tend to be significantly more metrically aligned than chance would predict.

The findings are summarized in figure 61. All percentages are based solely on unstressed, line-medial syllables. Each bar reflects the percentage of unstressed syllables in each position aligning with that position's preference (i.e. heavy in strong or light in weak). I separate the data into two charts, one (left) for weak positions (2, 4, 6) and the other (right) for strong positions (3, 5, 7). Each position anchors two bars, the first (dark) representing the actual corpus and the second (light) the prose model, representing chance expectations.[45]

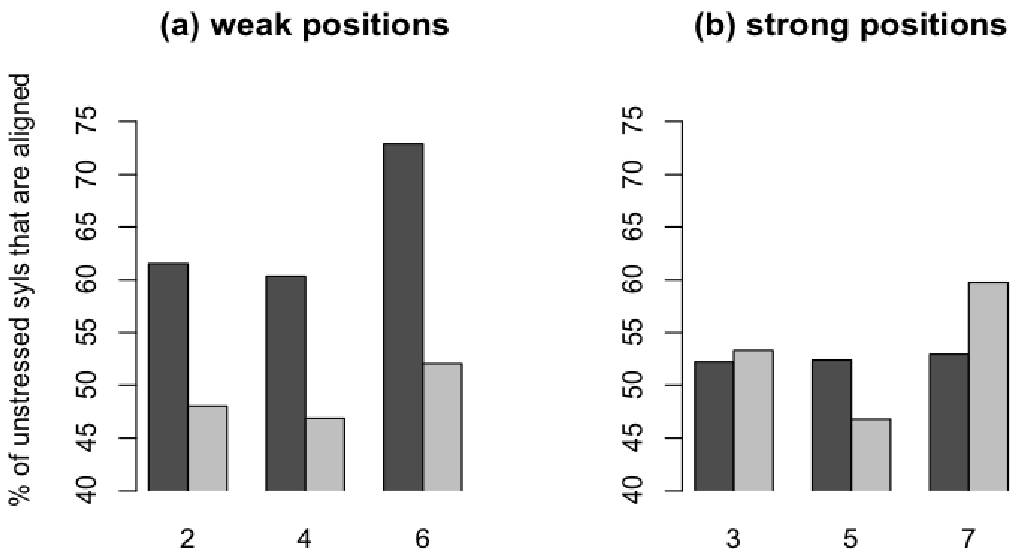

Figure 61: Alignment of unstressed syllables (dark = actual, light = control).

Among weak positions, the trend is clear. In every case, the actual corpus is significantly more aligned than the artificial one (highest [i.e. weakest] $p = 2 \times 10^{-104}$). This holds even for the weak of the first foot, which has been claimed to be metrically unregulated (Sadeniemi 1951, Kiparsky 1968:168). Most strikingly, in the final (non-anceps) weak, unstressed heavies are almost half as frequent as expected (27% observed heavy [i.e. non-aligned] vs. 48% expected). Thus, non-word-initial heavies are consistently preferentially avoided in weaks. Among strong positions, on the other hand, no consistent trend emerges. The difference in position 3 is nonsignificant

[45]Given the randomness of prose model construction, figures for the prose model come out slightly different from one run to the next. Nevertheless, given the sizes of the corpora and the impressive p values of the significant contrasts here, a reversal of a significant finding on any given run is extremely unlikely. For example, over 100 iterations of corpus construction, the mean per-position standard deviation was 0.4%. Moreover, checking minima and maxima reveals that no significant contrast was ever reversed in the 100 corpora (therefore, Monte Carlo $p < .01$ for each significant contrast; see §3.4.2 on the Monte Carlo method).

($p = .20$). Position 5 is significantly more aligned in the Kalevala ($p = 4 \times 10^{-12}$). However, position 7, the final strong in the line, bucks the trend, being significantly more aligned in the prose model ($p = 7 \times 10^{-30}$). I cannot explain this reversal in the cadence, but conclude that in general, and especially in weak positions, poets are sensitive to weight in non-initial syllables, many of which, after all, might receive secondary stress (though a possible role for secondary stress is not specifically tested here).[46] This conclusion is reinforced by logistic regression in the following section.

4.3.2 Intra-heavy weight in unstressed syllables

The rime V̆C is significantly lighter than Vː in non-initial positions in the word. This holds, in fact, not only for non-initial positions aggregately, but for each of the second and third positions taken independently. In both, V̆C patterns as significantly lighter (i.e. weak-skewed) than Vː, as shown in figure 62. Probabilities were computed by logistic regression controlling for word shape confounds as in §4.2.3. By the fourth syllable, as data become sparser, the contrast ceases to reach significance.[47]

syllable 1	V̆C $<$ Vː	$N = 13,080$	$p < .00001$
syllable 2	V̆C $<$ Vː	$N = 15,545$	$p < .00001$
syllable 3	V̆C $<$ Vː	$N = 8,196$	$p = .001$
syllable 4	V̆C $\sim$ Vː	$N = 2,006$	$p = .103$

Figure 62: Testing V̆C $<$ Vː across positions of the word.

Thus, not only is weight mapping the Kalevala meter sensitive to intra-heavy

[46]Summing over all positions, the actual corpus is more aligned ($p = 4 \times 10^{-165}$).

[47]Vː is rare in all non-initial positions ($N = 158$ in my corpus, excluding the first and last positions in the line), but not so much so in the second and third positions that it fails to achieve significance in figure 62.

weight, it is sensitive to intra-heavy weight in both stressed and unstressed positions. Note that while the first position in the word always receives primary stress in Finnish, the second position is said to be uniformly unstressed, not even receiving secondary stress, given clash avoidance (Kiparsky 2003:126).

Summarizing the Finnish findings, the most salient distinction in the metrics is that between heavy and light. At the same time, however, the poets exhibit significant sensitivity to an intra-heavy scale of weight comprising at least three significantly differently distributed levels (perhaps many more): $\breve{\text{V}}$C < VV < VVC (regardless of whether VV subsumes only long vowels, only diphthongs, or both long vowels and diphthongs). Moreover, unstressed syllables appear to be metrically regulated (though not as strongly as stressed ones), a novel result, as far as I am aware. Like stressed syllables, the regulation of unstressed syllables evidently extends into the intra-heavy realm, since even among unstressed syllables significant $\breve{\text{V}}$C < VV skews are observed. Although the present chapter is confined to the skeletal structure of the rime, it suffices to illustrate that weight is treated as an interval scale in Kalevala metrics, just as it is in the other traditions treated in this thesis: Not only are multiple tiers of weight evident within the heavies, but the differences between them are matters of varying degree, as opposed to strict separation.

5 Epic Sanskrit: šloka

5.1 Metrical and corpus preliminaries

The *šloka* [ɕloːkə] (also known as the *anuṣṭubh* [ʌnuʂʈup]) is the most common meter in Classical Sanskrit, known especially for its dominant role in the epics, though it is attested, with various changes, from the earliest Indo-Aryan literature (Oldenberg

1888, Arnold 1905, Macdonell 1916). Each *šloka* verse comprises two sixteen-syllable lines, with each line in turn comprising two eight-syllable half-lines (called *pāda*-s, literally, 'feet'). Anglophone scholars vary as to whether they use 'line' to refer to the sixteen- or eight-syllable unit. For convenience, I reserve 'line' for the sixteen-syllable half-verse. First, the orthographic line in epic manuscripts is usually two *pāda*-s. Second, sandhi can apply across half-line boundaries, but rarely if ever across sixteen-syllable units. Third, the metrical constraints on odd and even *pāda*-s differ. For example, the second *pāda* virtually always ends with a diiambic cadence $\smile - \smile \circ$ (light, heavy, light, anceps), whereas the most frequent cadence for the first is $\smile - - \circ$, with its final trochee creating a suspense, to be resolved with the regular diiambic cadence of the second half-line.

My Epic Sanskrit corpus comprises 229,118 *šloka* lines harvested from the two Sanskrit epics, the Mahābhārata (224,741 lines) and the Rāmāyaṇa (38,038 lines), as available online at the Göttingen Register of Electronic Texts in Indian Languages (`www.sub.uni-goettingen.de`, accessed c. 2005). These texts are not uniformly *šloka*, so I ran a script to retain only sixteen-syllable lines with the *šloka* cadence, $\smile - \smile \circ$ (87% of the original texts).

Because of its syllable count requirements, metrists classify the *šloka* as *akṣara-vṛtta*, i.e. syllabic, rather than *mātrā-vṛtta*, i.e. moraic, which would involve a regulated mora count but flexible syllable count (Velankar 1949, Allen 1973:61, Deo 2007, Fabb and Halle 2008:233). Nevertheless, it is still appropriate to speak of the *šloka* as being a quantitative meter, since in addition to its strict syllable count, syllable weight is regulated in some contexts. A syllable ending with a short vowel is light (*laghu*, cognate with *light*); all others are heavy (*guru*, cognate with *gravity*). Intervocalic CC is syllabified as V̌C.CV, even when the cluster could be a word onset (e.g. *tat.ra*, *tras.ta*). Recall that splitting such clusters is also the default in Homeric

Greek, though not as strictly observed there (§3.2). More generally, any nucleus followed by more than one consonant constitutes a heavy syllable; that is, at least if the nucleus is short, at least one consonant must be recruited from the following interlude to close the preceding syllable. Otherwise, following traditional accounts, onset maximization is employed (cf. Hermann 1923:257ff, Devine and Stephens 1994:41ff, Kessler 1998). For instance, *saṃskṛtam* 'Sanskrit' (ṃ = probably [n] in this case, ṛ = [ɽ̩]) is parsed as *saṃ.skṛ.tam*, not as *saṃs.kṛ.tam*. Finally, word boundaries within the line are ignored for basic scansion. Thus, *eva trayam* scans as *e.va‿t.ra.yam*, i.e., ——˘— (NB. Sanskrit 'e' and 'o' are always long; length is elsewhere indicated by a macron).

Some distributional restrictions in the *šloka* can be stated in a context-free manner. For example, positions 13 and 15 (counting from 1 to 16) must be light, position 14 must be heavy, and positions 1, 8, 9, and 16 are always anceps, as summarized in figure 63, in which Z is a placeholder for positions that I have not yet discussed. These context-free facts reflect two generalizations: First, half-line-peripheral syllables are always anceps. Second, lines must end with a diiambic cadence (modulo the previous license).

◦ZZZ ZZZ◦ ◦ZZZ ˘—˘◦

Figure 63: Context-free constraints in the *šloka*.

The term anceps is used (as above) for positions that could be either heavy or light without rendering the line unmetrical, but that is not to imply that there are no tendencies or preferences in those positions. Indeed, of the four anceps positions in figure 63, positions 1, 8, and 9 are all preferentially light in the aggregate, while

16 is preferentially heavy in the aggregate.[48]

The remaining (more or less) hard constraints can be expressed only contextually. I begin with the second half-line, as it is simpler to describe, being more constrained. First, positions 10 and 11 (the second and third in the *pāda*) cannot both be light. Second, positions 11 and 12 cannot comprise an iamb, as this would create an illicit triiambic cadence, cuing the cadence too early. A finite-state machine describing the set of metrical second half-lines is given in figure 64 (in which X = heavy or light).

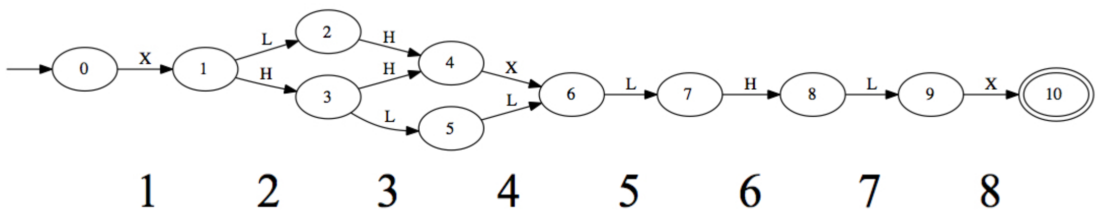

Figure 64: Licensed second half-lines in the *śloka*.

The first half-line is more complex. As with the second half-line, its second and third positions cannot both be light. Beyond this, the rules as traditionally stated, e.g., by Macdonell (1927:232) and Coulson (1992:250,310), are more arbitrary. Specifically, the first half-line is claimed to be confined to the partially specified templates in figure 65, given with their descending frequencies in the epics. Some of the five options in figure 65 are correlated with caesural constraints which I have

[48]As a rough assessment of these tendencies, the overall proportion of heavies in word-final syllables is 73.3%. But in position 16 it is 83.6% (significantly above chance) and in position 8 it is 68.8% (significantly below chance). Likewise, the overall incidence of heavies in word-initial syllables is 77.7%, whereas in the half-line-initial positions it is 59.4% and 61.6%, respectively. Nevertheless, a more accurate model of these observed vs. expected discrepancies would also control for word shape, not just word position.

omitted. A composite minimal finite-state representation follows in figure 66.[49]

(i)	XXXX	⏑––X	(86.9%)
(ii)	XXX–	⏑⏑⏑X	(5.0%)
(iii)	XX⏑–	–––X	(3.7%)
(iv)	XX⏑–	–⏑⏑X	(2.8%)
(v)	XXX–	–⏑–X	(1.1%)

Figure 65: Frequencies of *śloka* first half-line types in the epics.

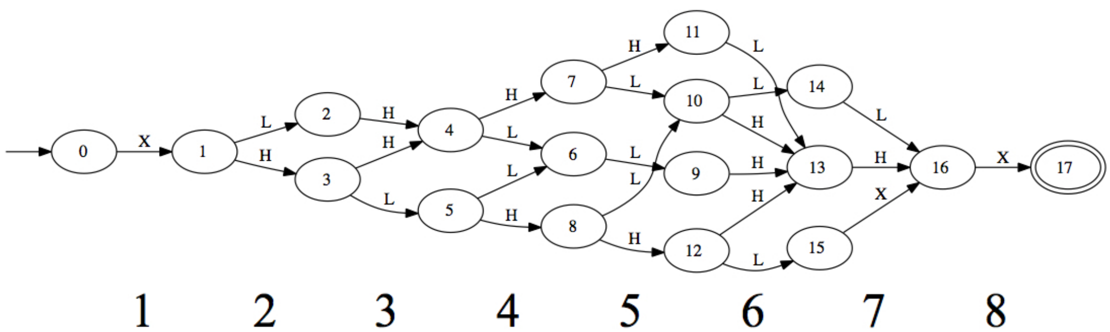

Figure 66: Licensed first half-lines in the *śloka*.

In my corpus, approximately 0.5% of lines fail to scan even by any of these options. While there are occasional transcriptional errors in the corpus (I did not hand-check these exceptions), it seems most likely that we are observing here a tapering off of increasingly marked options, rather than a strict cutoff in metricality (cf. the notion of complexity in Halle and Keyser 1971). Moreover, enumerating options in this manner fails to capture the generalizations underlying the meter. Nonetheless, the generative analysis of the *śloka* is beyond the scope of this paper. As discussed above when

[49]To generate the directed graphs in this section, I took the set of 2^{16} logically possible binary sequences and filtered out the ones that violated any of the constraints in Macdonell (1927) or Coulson (1992). I translated the remaining 1,120 templates into a finite-state machine with 1,120 nonbranching paths from start to end. I then used AT&T's FSM Library software (Mohri et al. 2009) to minimize the machine into the fewest states possible by running the `fsmdeterminize` and `fsmminimize` methods. (See also fn. 32.)

dealing with the unclear metrical situation in Tamil (§2.2), for the present purposes, it is necessary only to be able to gauge the heaviness propensities of metrical contexts, not to pin down every aspect of the templatic model.

It is crucial for the present enterprise to be cognizant that a position's status as regulated or free can often only be determined contextually. For example, whether position 2 is anceps or crucially heavy depends on its context. If position 3 is heavy, position 2 is anceps; otherwise, position 2 is crucially heavy. While certain positions, such as the *pāda* (half-line) peripheries, are always anceps, most ancipitia (edges labeled 'X' in the graphs above) are like the one in position 2 in that they can only be identified as such in the context of their line.

5.2 Intra-heavy weight in Sanskrit

I therefore apply the same context-sensitive LINEAR PROPENSITY model employed for Tamil in §2.5.1 to the Sanskrit corpus. As in Tamil, the propensity of each syllable's position is gauged by collecting all lines of the same length (a vacuous restriction in the present case, as every line in the Sanskrit corpus is sixteen syllables) and exhibiting the same five-syllable window centered on the position in question (e.g. HL_LH in position 4). The log ratio of heavies to lights in that position in that subset of lines is then taken as the estimate of the weight propensity of the position, with positive values indicating preferentially heavy positions and negative values preferentially light ones. For more details of implementation, see §2.5.1. In the present case, each syllable token exhibits on average 21,068 comparanda, and a total of 2,106,214 heavy tokens are used as data points (it is unnecessary to also include lights in the data, since it is uncontroversial that they are lighter than heavies).

Figure 67 is the regression table for four levels of skeletal rime structure (as above): $\breve{\mathrm{V}}$T (where T = any obstruent, including the letter *visarga*, i.e. [h]), $\breve{\mathrm{V}}$N

(N = sonorant, including the letter *anusvāra*, a chameleonic nasal), VV (as always covering both long vowels — including orthographic *e* and *o* — and diphthongs, which also scan as long), and finally VVC. No significant difference is found between VV and VVC, but otherwise the hierarchy is consistent with those inferred from all the previous case studies: V̆T < V̆N < VV, VVC. A Hasse diagram follows in figure 68.

rime		coefficient	standard error	*z*-value	*p*-value
intercept	(=V̆T)	1.8940	.0312	60.6	$< .00001$
V̆N	[vs. V̆T]	.0438	.0027	16.5	$< .00001$
VV	[vs. V̆N]	.0311	.0024	13.0	$< .00001$
VVC	[vs. VV]	−.0001	.0028	−.0	$= .98$

Figure 67: Linear regression model for skeletal rime structure in Sanskrit.

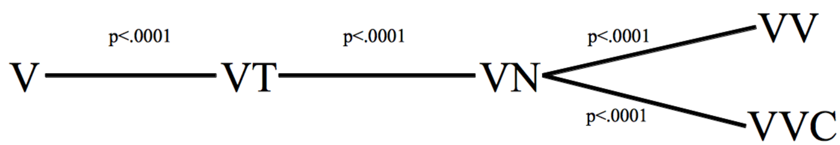

Figure 68: Hasse diagram for Sanskrit rime skeletons.

In conclusion, Epic Sanskrit, like every other quantitative meter examined here, is evidently sensitive to (at least a few and possibly many more) intra-heavy grades of weight, as diagnosed by distributional skews of heavy syllable types between preferentially lighter vs. heavier positions. For example, the heavier the position, the more overrepresented (relative to lighter positions) VVC_0 becomes relative to lighter types such as V̆N and V̆T. This subcategorical hierarchy aligns not only with those of the other meters examined but also with the crosslinguistic typologies of other weight diagnostics (see §7).

6 Old Norse: skaldic dróttkvætt

6.1 Metrical and corpus preliminaries

Old Norse poetry (c. 700–1300 CE) comprises two genres, Eddic and skaldic. I focus on the latter here, particularly the Old Icelandic *dróttkvætt*, the most widely attested skaldic meter. Modal line length is six syllables, though lines can be longer due to the option of filling certain positions with two syllables. Every line ends with a heavy stressed syllable followed by an unstressed syllable. Because stress is almost uniformly word-initial (unstressed prefixes being relatively marginal; Russom 1998:13ff), it follows that the line typically ends with a disyllabic, heavy-initial content word.

The metrical description of the preceding four positions is more vexed. Perhaps the simplest proposal is that of Craigie (1900:381), who proposes two metrical templates for the *dróttkvætt*, SWSWSW (i.e. trochaic trimeter) and SSWWSW (adding an inversion, though cf. Getty 1998 for arguments that inversion does not necessarily implicate multiple templates). To this scheme, Árnason (1991:124ff, 1998:102) adds a third template, WSSWSW. A more elaborate and also more widely employed description ('still the model most commonly referred to by philologists' –Árnason 1998) is Sievers' (1893) five-type taxonomy for Germanic verse; cf. Kuhn (1983) and Gade (1995) for revisions in this tradition. For the sake of illustration, I employ Árnason's (1991) system, as exemplified in figure 69. In this scheme, a strong position can only be filled by a stressed, heavy syllable, whereas weak positions are not as strictly regulated.

(a)	*gróðr sá fylkir fáði*					
	S	W	S	W	S	W
	gróːðr	sáː	fýl	kir	fáː	ði
(b)	*ungr stillir sá milli*					
	S	S	W	W	S	W
	úngr	stíl	lir	sáː	míl	li
(c)	*svartskyggð bitu seggi*					
	S	S	W	W	S	W
	svárt	skyggð	bí	tu	ség	gi

Figure 69: Three *dróttkvætt* lines scanned.

Because syllabification is also a vexed issue in Norse metrics, I consider two very distinct approaches, not with a view to arguing for one or the other, nor to suggest that the correct algorithm is necessarily either, but merely to show that even with two opposing extremes (and, by hypothesis, any more nuanced intermediate position), the same general trend in gradient weight is observed. At one extreme, onset maximization (OM; as in figure 69) prioritizes building onsets that are as complex as the phonotactics permits, e.g. [hun.drað] (cf. Árnason 1991:123 for a qualified version of this approach). At the other extreme, coda maximization (CM) groups all consonants with the preceding vowel, if any, e.g. [hundr.að] (Hoffory 1889:91, Beckman 1899:68, Pipping 1903:1, 1937, Kuhn 1983:53, Gade 1995:30).[50] Note the asymmetry between these approaches: While OM is reined in by phonotactics ([hun.drað], not *[hu.ndrað]), CM (in the tradition cited) is not.[51] The criterion for light vs. heavy depends on the algorithm. Under OM, the rime $\breve{V}$ alone is light. Under CM, $\breve{V}$, $\breve{V}$C, and VV are light.

[50]In an independent vein of research, Steriade supports the same algorithm, terming the spans 'intervals' rather than 'syllables' (Steriade 2008b, 2009, 2011, cf. Steriade 2008a).

[51]As in Finnish, the parser here does not resyllabify across words (Gade 1995:31), though this issue also deserves more scrutiny in Old Norse.

6.2 Logistic model: stresed syllables

A corpus of 11,832 six-syllable *dróttkvætt* lines was harvested from the University of Sydney Skaldic Project (`skaldic.arts.usyd.edu.au`, accessed August 2010). Though the *dróttkvætt* is not confined to six syllables, retaining only six-syllable lines facilitates metrical parsing. Under Árnason's scheme, position 5 is always strong and positions 4 and 6 are always weak.[52] Additionally, position 3 is weak if and only if positions 1 and 2 are both strong (filled by stressed heavies). Because positions 1 and 2 are more variable (being SS, SW, or WS), I put them aside here. Each syllable from the final four positions is coded for the skeletal structure of its rime (e.g. $\breve{\mathrm{V}}$C), its position type (1 for strong, 0 for weak), and its word context (as in §4.2.3). A logistic model then predicts metrical placement from rime type, factoring out word context as a random effect as before.

Figure 70 is the resulting regression table, at this point considering only stressed, word-initial syllables, as in Finnish (§4.2), and assuming OM. As always, the table is forward-difference coded, such that a positive coefficient indicates that the given rime type exhibits greater bias towards strong positions than the comparandum type. The hierarchy is thus $\breve{\mathrm{V}} < \breve{\mathrm{V}}\mathrm{C} < \mathrm{VV} < \mathrm{VVC} < \breve{\mathrm{V}}\mathrm{CC} < \mathrm{VVCC}$ (every link $p \leq .0001$).

[52] It is not a consensus that the fourth position is always weak; see Árnason (1991:139, 2009:48) for discussion of the issues.

rime	comparandum	coefficient	standard error	z-value	p-value
intercept	(i.e. V̆)	−.7075	.7082	−1.00	= .32
V̆C	(vs. V̆)	.3951	.0561	7.04	< .00001
VV	(vs. V̆C)	.2098	.0482	4.35	= .00001
VVC	(vs. VV)	.2800	.0734	3.81	= .0001
VCC	(vs. VVC)	.4788	.0752	6.36	< .00001
VVCC	(vs. VCC)	.4397	.0902	4.87	< .00001

Figure 70: Logistic model for Old Norse stressed syllable placement.

The same basic hierarchy is observed (every link $p < .0001$) if CM is instead employed: V̆C < V̆CC < VVC < VVCC < V̆CCC < VVCCC (every link $p < .0001$). An additional consonant is now appended to each rime to better align the OM and CM scales. For example, the initial rime of *rifu* is V̆ under OM but V̆C under CM.[53] Figure 71 compares the two hierarchies graphically. To facilitate comparison, the intercepts are normalized to zero. To visualize the global trend, the forward-difference coded coefficients are presented cumulatively (i.e. as sums of coefficients up to and including the given rime). Rimes on the x-axis are labeled according to both schemes, with OM on top. While it is not surprising that these scales are well correlated, this comparison demonstrates that the choice of syllabification algorithm does not qualitatively alter the conclusion concerning weight.

[53]This is not to imply that there is a biunique mapping between OM and CM rimes, in which case they would be notational variants. Recall *hundrað*, in which the initial rime is V̆C under OM and V̆CCC (not V̆CC) under CM. It merely reflects that V̆ under OM is most frequently V̆C under CM, and so forth.

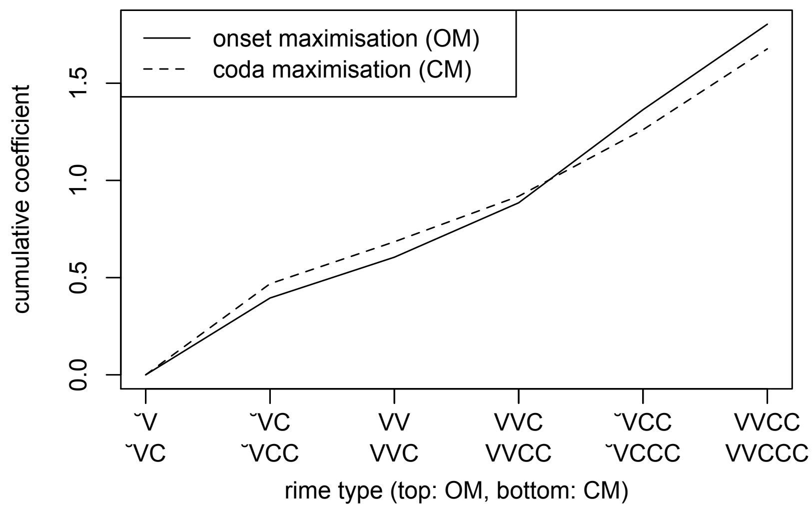

Figure 71: Weight coefficients under two syllabification algorithms.

6.3 Logistic model: unstressed syllables

I now turn to unstressed (non-word-initial) syllables. Because they cannot occupy strong positions, strong/weak asymmetries cannot be used as a diagnostic, as they were above for stressed syllables. Instead, I capitalize on the increasing rigidity of the meter towards the end of the line. In particular, I compare unstressed syllables in positions 4 and 6, which are the final two weak positions and also the only two positions that are uniformly weak, to those in all other positions. The former are coded 0 and the latter 1, reflecting the hypothesis that unstressed syllables in uniformly weak cadential positions will tend to be the aggregately lighter set. The logistic model is otherwise set up as above.

Under OM, the following hierarchy emerges: V̆ < V̆C < VV < V̆CC < VVC < VVCC (every link $p < .0002$). Under CM, the same hierarchy emerges (every link $p < .0001$, except V̆CC < VVC, which is $p = .003$). The coefficients under both schemes are plotted in figure 72, as they were in figure 71.

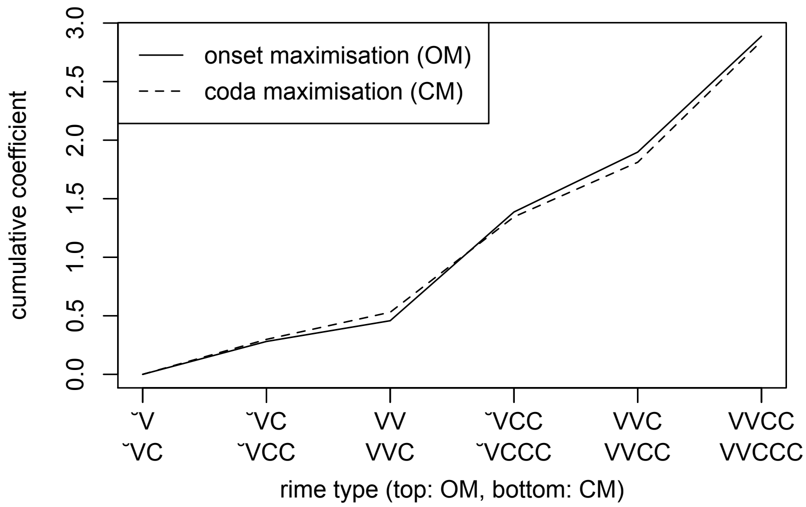

Figure 72: Weight coefficients under two syllabification algorithms.

6.4 Old Norse syllable weight: synthesis

In conclusion, two tests reveal sensitivity to a scale of weight in *dróttkvætt* composition. First, the heavier a stressed syllable is, the more likely it is to be placed in a strong position, revealing the scale (in OM terms) V̆ < V̆C < VV < VVC < V̆CC < VVCC. Second, the heavier an unstressed syllable is, the less likely it is to be placed in a cadential weak position, revealing the scale V̆ < V̆C < VV < V̆CC < VVC < VVCC. These tests are independent of each other, relying on completely disjoint sets of data, yet reveal tightly correlated hierarchies. The one exception concerns the rimes V̆CC and VVC, which are adjacent under both tests, but in opposite orders. The composite hierarchy is therefore V̆ < V̆C < VV < {V̆CC, VVC} < VVCC (where the status of the braced pair is unclear), as in figure 73, consistent with the V̆ < V̆C < VV < VVC scales found in the preceding case studies.

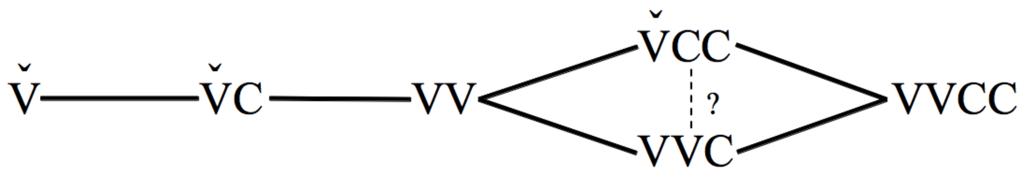

Figure 73: Hasse diagram for Old Norse rime skeletons.

Moreover, the present tests cast doubt on whether the *dróttkvætt* privileges any single binary criterion over the various other weight distinctions in the same way that Ancient Greek and Finnish appear to (see §10.1 below).

Part II
The phonetic interface of gradient weight mapping

7 Motivating gradient weight in Tamil

Figure 74 repeats from §2.5.4 an interval scale of syllable weight for Tamil metrics inferred from distributional asymmetries in Kamban̲'s epic. While §2 was dedicated to the descriptive tasks of extracting this type of scale from the corpus in a controlled manner and demonstrating its place as a productive factor in versification, little effort was made to motivate the particular features of the scale, relating them to functional (e.g. phonetic) and/or formal linguistic principles. It is to these issues of explanation that I now turn.

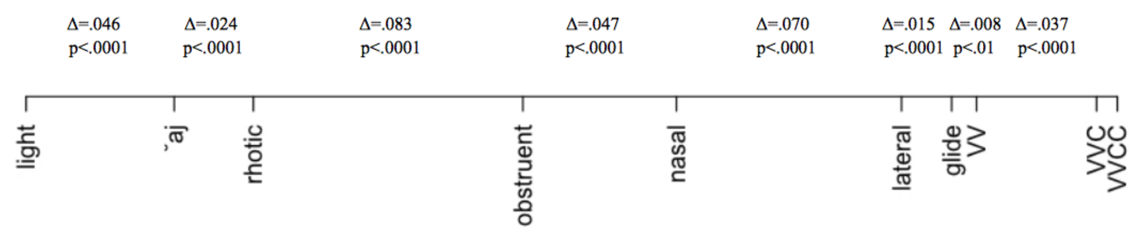

Figure 74: Estimated weights of ten rime types in Tamil.

One basic tentative principle is that more structure, e.g. complexity in segments or timing slots, tends to correlate with (if anything) greater weight (Gordon 2002). In the present case, if we ignore features of segments, including the vowel-consonant distinction, and observe only the number of timing slots in the rime, the progression is monotonic: X < XX < XXX (<) XXXX (where '$x < y$' is to be read 'x is lighter than *y*' and XXXX is numerically but not significantly heavier than XXX).[54] Regardless of the qualities of the segments/slots involved, light syllables can be characterized as the set of syllables with simple, non-branching rimes, whereas branching rimes are (progressively) heavier (under moraic theory [Hyman 1985, Zec 1988, Hayes 1989a, Steriade 1991], for instance, it might be said that each rimal segment, whether vowel or consonant, contributes a mora; cf. Hayes 1979:196 on an X $<$ XX $<$ XXX hierarchy in Persian meter).

It is also obvious, given the number of distinctions in figure 74 (and possible additional distinctions that are not shown), that segmental complexity is not the whole story. A second commonly invoked principle of weight — one sometimes thought to be universal and as such hardwired into theoretical proposals (e.g. Zec 1995, 2003) — is that greater sonority is correlated with greater weight (cf. e.g. Zec 1988, 1995, 2003, Morén 1999, Gordon 2006, de Lacy 2002, 2004). For example, in some languages, the weight of $C_0\breve{V}C$ depends on the features of the coda. Conventional wisdom holds that such a distinction will coincide with a sonority cutoff, with the more sonorous subset being the heavier. In Kwakwala, for one, $C_0\breve{V}C$ is heavy if and only if the coda consonant is a sonorant (Boas 1947, Zec 1995). A more common weight distinction is rimal $\breve{V}C < VV$,[55] as also seen in figure 74, which can likewise be explained by the greater sonority of the latter, even if the two categories are comparable in both duration (as they often are) and segmental complexity (if VV is a diphthong).

A crosslinguistically typical sonority hierarchy is given on the top of figure 75

---

[54]Perhaps the only category with an ambiguous number of segments is the light diphthong [ăj]. The stated generalization remains unaltered, however, regardless of whether [ăj] is considered one or two segments.

[55]This is almost always the polarity of this distinction if one is made, but a handful of perhaps controversial exceptions can be cited (see, e.g., references in Devine and Stephens 1994:72).

(Hogg and McCully 1987:33, Parker 2002), ranging from least to most sonorous. For comparison, the Tamil weight scale from figure 74 is given on the bottom of figure 75, with association lines indicating the alignment between the two scales. The only Tamil categories from figure 74 that are excluded from figure 75 are the light diphthong [ăj], to be treated in a moment, and the superheavy rimes VVC and VVCC, which are off the scale, but entirely consistent with it (particularly if any C is considered to be more sonorous than Ø, as figure 75 implies). As regards the crosslinguistic sonority hierarchy, obstruent can sometimes be further divided, e.g. voiced stops sometimes pattern as (if anything) more sonorous than voiceless ones, or fricatives as (if anything) more sonorous than stops in general (Hogg and McCully 1987). However, these distinctions, which are at any rate uncommon, are moot for (Middle) Tamil, which lacks fricatives and (phonemic) voicing altogether; a coda obstruent is always a voiceless stop in conservative Tamil.[56]

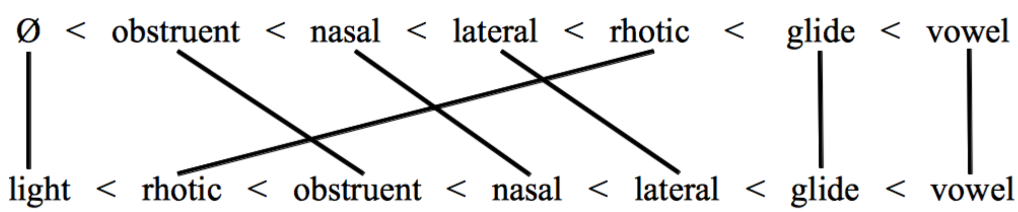

Figure 75: Typical sonority scale (top) vs. Tamil metrics (bottom).

The two categories in the Tamil scale in figure 74 that are not readily explained by either of these two structural principles — segmental complexity and sonority — are the diphthong [ăj], which is ostensibly both segmentally complex and highly sonorous throughout, being glide-final, and V̌R (R = rhotic), which is expected to

[56]One exception is the rare letter *āytam*, probably a fricative in Kamban̲'s time (Ryan 2003, Krishnamurti 2003). This sound is not included in any of the categories here.

be intermediate in weight between V̆L (L = lateral) and V̆W (W = glide), given the sonority scale at the top of figure 75.

7.1 The lightness of the diphthong [ăj]

As discussed in §2.1 (especially fn. 6), the diphthongs [ăj] and [ăʋ] are traditionally treated as categorically light in Tamil, at least in non-initial position. In initial position, they pattern as bimoraic and are approximately as long as long vowels. (Because [ăj] is hundreds of times more frequent than [ăʋ], for the purposes of exposition, I focus on [ăj].) While these are falling-sonority diphthongs and it is more common for rising-sonority diphthongs to be classified as light (e.g. McCarthy 2000:152, with references), languages/processes treating at least certain falling-sonority diphthongs as light, or as an intermediate grade between light and heavy, are amply attested, including Maori (Bauer 1993, Harlow 2001), Kara (de Lacy 1997), Gere (Paradis 1997:532), Tohono O'odham (Miyashita 2002), Finnish (Keyser and Kiparsky 1984, Kiparsky 2003), and perhaps English (Harris 1994:278).

The Tohono O'odham case is particularly reminiscent of Tamil, since its diphthongs have been claimed to be heavy word-initially and light elsewhere (Miyashita 2002). Finnish is similar to Tamil in a different way: On Kiparsky's (2003) analysis, stressed diphthongs are bimoraic and unstressed diphthongs monomoraic. Thus, a diphthong in the initial syllable (which receives primary stress) is always bimoraic, whereas one in the second syllable (which never receives stress) is always monomoraic. Recall that in Tamil, like Finnish, accent is arguably always word-initial (Keane 2003, 2006, Krishnamurti 2003).

Cursory phonetic analysis reveals Tamil [ăj] to be much closer to V̆ than to Vː. In figure 76, I give average durations in milliseconds (ms) for light (i.e. V̆) rimes, [ăj] rimes, and heavy rimes, respectively, in a recording of a high register (*cen-tamil̲*) of contemporary Tamil (as spoken by Kausalya Hart in the audio materials accompanying Hart 1999). Measurements here are exclusively from word-medial syllables. The corresponding boxplot is given to the right. (More rigorous analysis of these correlations using more data is pursued in §9.)

rime type	mean duration (ms)	N
light	71.8	48
[ăj]	89.3	5
heavy	184.2	79

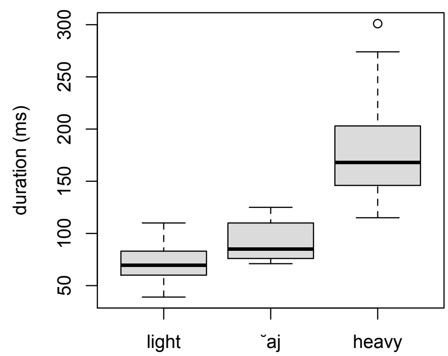

Figure 76: Duration of [ăj] relative to light and heavy.

An amplitude waveform (top) and spectrogram (bottom) of one (93 ms) token of [ăj] from the word [manăjʊij-um] ‘wife-and’ is given as an example in figure 77 (made with Praat; Boersma and Weenink 2011). In this spectrogram, the vowel transcribed as [ăj] is pronounced closer to [ĕj] by the present-day speaker (see fn. 6), but its transcription is standardized according to its more conservative form.

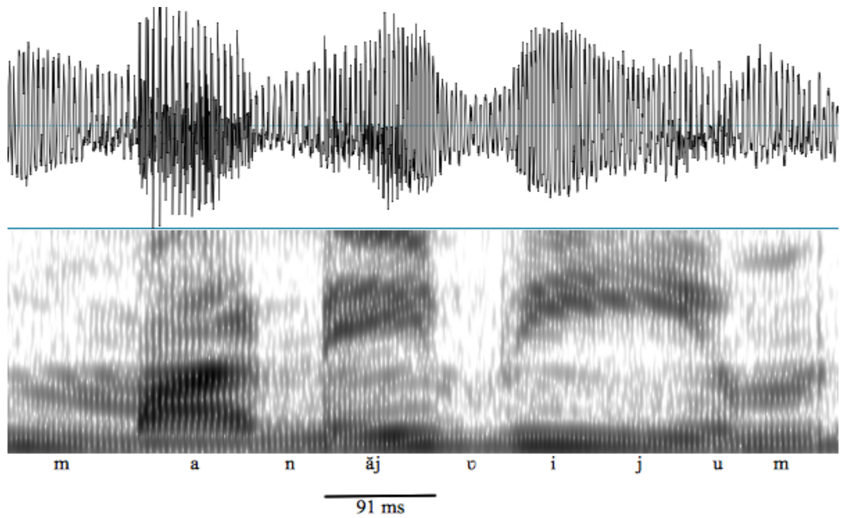

Figure 77: Waveform and spectrogram of [ăj] in <*maṉaiviyum*>.

Despite the eight centuries of separation between Kambaṉ and the recording, the phonetic measurements align uncannily with the weights estimated purely from metrical corpus analysis.[57] Figure 78 recapitulates (on the bottom side of the continuum) the metrically-derived scale in figure 74, showing only the categories light, [ăj], and heavy (now averaged over all heavies). On this continuum, [ăj] is 17.2% of the way from light to heavy. In the durational data in figure 76, [ăj] is 15.5% of the way from light to heavy; this difference is shown on the top side of the continuum in figure 78. The relative position of [ăj] between lights and heavies in the two independent continua is almost identical. (The position of [ăj] is all that is of interest in figure 78; the light and heavy endpoints are rescaled to align with each other.)

[57]Using a measure of phonetic weight incorporating perceptual energy (Gordon 2002, 2005, Gordon et al. 2008) does not alter this conclusion.

Figure 78: Alignment of [ǎj] between phonetics (top) and metrics (bottom).

In conclusion, despite [ǎj] being highly sonorous throughout and (arguably) bisegmental, at least in earlier Tamil, the motivation for its treatment as relatively light — but still not equivalent to a short vowel — by the metrics is obvious when one considers its phonetics.

7.2 The peculiar lightness of Tamil rhotics

The relative lightness of the Tamil rime V̌R (R = any rhotic, including [ɾ] and [ɻ] in Middle Tamil) will receive a more detailed treatment here, both because it violates what is often assumed to be a linguistic universal concerning the correlation of coda sonority and weight (§7) and also because its special status has not, as far as I am aware, been previously reported. Despite being an uncontroversially bisegmental vowel-consonant sequence in which the coda consonant is highly sonorous (as will be reinforced in §8.3), V̌R is treated as lighter than other V̌C rimes, even those in which the coda consonant is less sonorous than the rhotics according to convergent phonetic and phonological criteria.

Moreover, the metrics is not the only phonological system diagnosing V̌R as lighter than all other V̌C; prosodic minimality independently supports the Tamil rhotics as being lighter than all other consonants, regardless of sonority. I begin with a brief discussion of phonetics in this section, as in §7.1, followed by an analysis of prosodic minimality in §8.

Figure 79 repeats the three rows from figure 76, adding a fourth row for the rime V̆R in word-medial position (all five tokens are [aɾ], by far the most frequent of V̆R syllables word-medially). Judging by duration, the conclusion once again parallels the one derived exclusively from the distribution of syllables in meter: V̆R tends to be closer to lights than heavies, though somewhat heavier than [ăj].

rime type	mean duration (ms)	*N*
light	71.8	48
[ăj]	89.3	5
V̆R	109.5	5
heavy	184.2	79

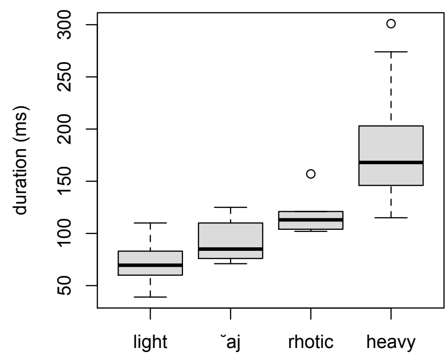

Figure 79: Duration of V̆R relative to light, [ăj], and heavy.

A waveform and spectrogram for one of the medial V̆R tokens is given in figure 80, followed by a scalar representation of the four categories (cf. figure 78) in figure 81. As figure 80 implies, any release of the tap (a short schwa-like svarabhakti vowel) is included in its duration measure. In this token, an underlying /ʋaɾk/ sequence is realized as closer to [ʋ^{ə}ɾ̩^{ə}h], with the tap being realized almost medially within a short [a]- or [ə]-colored nucleus (cf. rhotic-vowel metathesis, as in Steriade 1990, Blevins and Garrett 2004).[58] In more careful speech, however, the tap clearly follows the vowel and the vowel-svarabhakti ratio is larger (see, e.g., figure 85 for [ɾ] in final position). Moreover, vowel quality is contrastive before coda V̆R (e.g. [iɾ] and [aɾ]

[58]A case of productively optional alignment of V̆ and R can be found in Malinaltepec Tlapanec (Suárez 1983).

contrast as rimes).

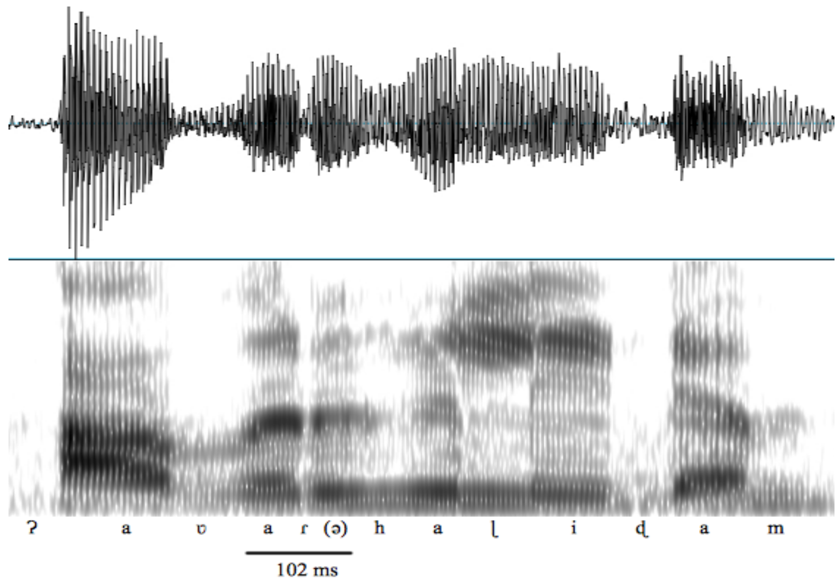

Figure 80: Waveform and spectrogram of [aɾ] in <*avarkaḷiṭam*>.

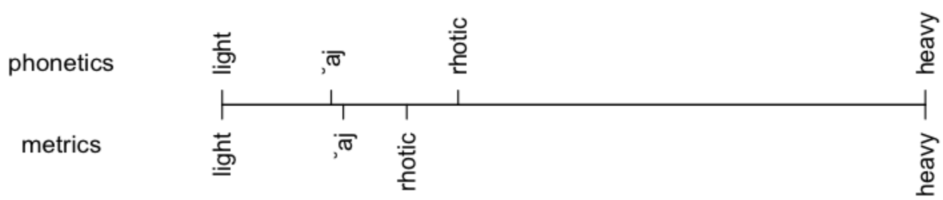

Figure 81: Alignment of [ăj] and V̌R between phonetics (top) and metrics (bottom).

8 Light rhotics: convergence between metrics and minimality

8.1 Rhotics do not contribute to minimality in Tamil

As in Latin (Mester 1994 and references therein), Tamil words of all types can be described as being minimally bimoraic, i.e., C_0VC (with a caveat below), C_0VV, or larger. In Latin, for instance, the root /da/ (as inferred from the infinitive *dă-re* 'to give', cf. *staː-re* 'to stand') is lengthened when it surfaces unaffixed for imperative singular *daː* 'give!', **da* (cf. *staː* 'stand!'). A standard prosodic analysis of this lengthening (Prince and Smolensky 1993/2004, Blumenfeld 2010; but see Garrett 1999) can be summarized as (a) grammatical words must be prosodic words, (b) prosodic words must dominate one or more feet, and (c) feet must be binary, dominating two moras (or two syllables).

Though this minimum is usually met by default in both Latin and Tamil, given that almost all actual roots are underlyingly bimoraic or larger (a correlation termed CONCURRENCE in Ketner 2006), subminimal roots exist in both languages, their treatment revealing that the grammar actively enforces minimality. In Tamil, the roots /ʋa/ 'come' and /t̪a/ 'give' are monomoraic (cf. infinitives *ʋa-ɾ-a* and *t̪a-ɾ-a*); both are lengthened when unaffixed, as imperative singulars: *ʋaː* 'come!', *t̪aː* 'give!' (**ʋa*, **t̪a*), as confirmed by both orthography and metrical scansion. Figure 82 underlines this point of parallelism between Latin and Tamil.

language	root	gloss	infinitive	imperative.2s
Latin	/staː/	‘stand’	staː-re	staː
Latin	/da/	‘give’	da-re	$\boxed{\text{daː}}$
Tamil	/puː/	‘flower’	puː-kk-a	puː
Tamil	/ʋa/	‘come’	ʋa-ɾ-a	$\boxed{\text{ʋaː}}$
Tamil	/t̪a/	‘give’	t̪a-ɾ-a	$\boxed{\text{t̪aː}}$

Figure 82: Repairing subminimality in Latin and Tamil.

Tamil differs from Latin in that (traditionally, at least) an isolated prosodic word cannot end with an obstruent.[59] Thus, $*C_0VT$ (T = any obstruent) is illicit in Tamil. This gap is motivated by phonotactics, not prosody/minimality. Even when minimality is not at stake, an isolated word cannot end with an obstruent (e.g. $*C_0VVT$, which is a licit syllable nonfinally). Moreover, even while the bimoraic minimum continues to be enforced in the modern language, obstruent-final words are beginning to enter the language (see also §8.2), corroborating the phonotactic nature of this gap.

There is, however, an exception to the aforementioned Tamil minimality generalization that cannot be attributed to phonotactics: $C_0\breve{V}C$ is minimal, but only if the coda is nonrhotic. If the coda is one of the two rhotics (ɾ and ɻ, Narayanan et al. 1999), the word is subminimal. There is a clear gap for $C_0\breve{V}R$ roots/words (R = any rhotic). Some monosyllabic words are exemplified in figure 83 (University of Madras *Tamil Lexicon*, 1924–1936).

[59] I say ‘isolated’ because an obstruent-final word can arise through sandhi with a following word.

short vowel		long vowel	
pon	‘gold’	poːn	‘trap’
poj	‘lie’	poːj	‘went (converb)’
*po	(subminimal)	poː	‘go’
*poɾ	(subminimal)	poːɾ	‘wear’
*poɻ	(subminimal)	poːɻ	‘be cleft’
kal	‘stone’	kaːl	‘leg’
kaj	‘hand’	kaːj	‘unripe fruit’
kaɳ	‘eye’	kaːɳ	‘sight’
*ka	(subminimal)	kaː	‘protect’
*kaɾ	(subminimal)	kaːɾ	‘be pungent’
*kaɻ	(subminimal)	kaːɻ	‘solidity’

Figure 83: Examples of monosyllabic words (and gaps).

The gap is entirely systematic ($p < .0001$), as figure 84 shows, which illustrates both token and type counts for the two rhotics in various phonological contexts in Kamban̲'s epic. The gap is also explicitly acknowledged by the earliest (c. 200 CE) indigenous Tamil grammar, the *Tolkāppiyam* (§1.2.14ff, Murugan 2000).

	V̆ɾ	V̆ɻ	VVɾ	VVɻ
final in monosyllable	0 (0)	0 (0)	2,356 (41)	269 (16)
final in polysyllable	10,617 (2,549)	345 (71)	4,954 (2,165)	43 (39)
word-medial	35,344 (14,199)	9,877 (4,195)	6,595 (3,043)	2,159 (838)

Figure 84: Token (type) counts of rhotics in various phonotactic contexts.

As figure 84 also clarifies, rhotic codas are felicitous after short vowels when minimality is not at stake; consider common words such as *avaɾ* ‘he (respectful)’, *t̪amiɻ* ‘Tamil’, and *kaɾvam* ‘pride’. When minimality is unthreatened, vowel length is contrastive before rhotics (e.g. *avaɾ* ‘he’ vs. *kiɻaːɾ* ‘water lift’; *kaɾmam* ‘action’ vs. *aːrmăj* ‘sharpness’).[60]

8.2 Loanword phonology and minimality

Loanword phonology, for its part, also supports $C_0\breve{V}R$ failing to achieve minimality. Loanwords in Tamil of the shape C_0VR almost invariably have a long vowel (as figure 84 suggests), even when the corresponding vowel in the donor language is short.[61] For example, Sanskrit words such as *sphira-* ‘abundant flow’, *dharā-* ‘house’, and *dur-* ‘bad’ (a prefix in Sanskrit) correspond to Tamil *piːr*, *t̪aːr*, and *t̪uːr*, respectively. Greek *ắrēs* ‘Ares’ was borrowed into Tamil as *aːr* ‘Mars’. In nonrhotic contexts, comparable lengthening is not found (e.g. *kam* and *karmam* ‘act’ < Sanskrit *karma-* or Prakrit *kamma-*).

Although English loanwords are numerous in contemporary Tamil, they are arguably not diagnostic of minimality. They are, to be sure, consistent with $C_0\breve{V}R$ being subminimal. English words such as *sir* and *car* are invariably borrowed with long vowels (*saːr* and *kaːr*), while other consonant-final monosyllables with lax vowels are borrowed with short vowels, e.g. *kap* ‘cup’, *cek* ‘check’, *mes* ‘mess’, and *pen* ‘pen’ (Hart 1999; as in Japanese, lax vowels tend to map to short vowels; Takagi and Mann 1994, Dupoux et al. 1998:11).

On the one hand, this discrepancy is consistent with hypothetical words such as

---

[60] I confirmed in the recordings that the initial vowel of words such as *karvam* is indeed pronounced as short, averaging 134 ms for three relevant tokens, roughly half as long as the long vowel in initial (C)V:R(CV...). Thus, vowel length is contrastive before coda R even in word-initial position when minimality is not at stake.

[61] The only exception that I have encountered is the brief unsourced entry *ṭar* in the *Madras Tamil Lexicon* for Hindi *ḍar* ‘fear’, though this word has no entry (with a long or short vowel) in a contemporary Tamil dictionary, *Kariyāviṉ Taṟkālat Tamiḻ Akarāti* (Cre-A, Madras: 1992) and was absent from my poetic corpora. It is possible in this sporadic case that the orthography was rendered faithfully to the Hindi irrespective of the typical Tamil pronunciation.

**saɾ* and **kaɾ* being subminimal and therefore repaired by lengthening. On the other hand, words such as *sir* and *car* are pronounced with long vowels in British English (e.g. *sɜː* and k^h*aː*, though the British dialect from which these words were originally borrowed might have been rhotic, given that Tamil borrowed them with taps). As further support for this second possibility, Japanese has borrowed these same two words with long vowels (*saː* and *kaː*) even though it ostensibly lacks a bimoraic minimum (cf. *e* 'picture').[62] The loanword *paːɾk* 'park' further supports that stressed English *-ar-* is simply rendered as *aːɾ* in Tamil irrespective of minimality (since *paɾk* would be minimal).[63] Thus, English loanword phonology is mute on the question of whether $C_0\breve{V}R$ is subminimal or not, while loanwords from other languages, such as Sanskrit, support this conclusion.

8.3 The Tamil rhotics are highly sonorous consonants

In this section, I intend to clarify two points, first, that the Tamil rhotics are in fact true consonants (rather than vowel colorations), and second, that they are highly sonorous ones at that, intermediate between the laterals and glides in sonority, as would be expected on crosslinguistic grounds. Conservative Tamil, like Malayalam (Asher and Kumari 1997) and arguably Proto-Dravidian (Krishnamurti 2003), distinguishes two rhotics, namely, the prealveolar tap ɾ and the palatal rhotic approximant

---

[62]Short vowels in Japanese monosyllables uttered in isolation undergo significantly more lengthening than other final short vowels in phrase-final position, indicating that perhaps a prosodic minimum is enforced in some sense, though, at the same time, the lengthened short vowels of isolated monosyllables are still shorter than the realizations of underlying long vowels (Mori 2002, Kawahara 2011).

[63]The spelling *paɾk* (in Tamil script) is attested several thousand times on Google (`www.google.com`, accessed November 2010), but still only about 2% as often as the prescribed long vowel spelling. Another consideration is that *park* is also borrowed with a long vowel into other Indian languages such as Hindi, possibly influencing the Tamil convention.

ɻ (see Narayanan et al. 1999 for phonetic analysis). Segmented waveforms and spectrograms of the two rhotics are depicted in figures 85 (careful token of [aʋaɾ] ‘he’) and 86 (careful token of [t̪amiɻ] ‘Tamil’).

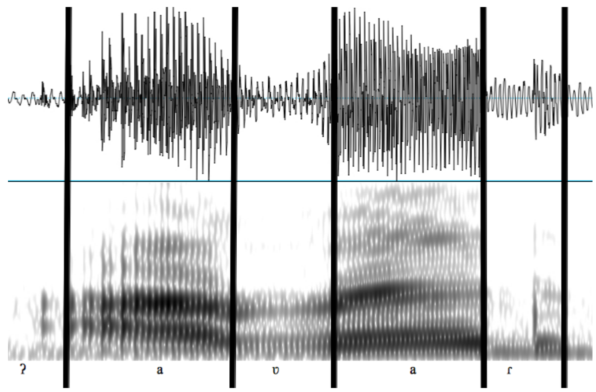

Figure 85: Waveform and spectrogram of [ɾ] in <*avar*>.

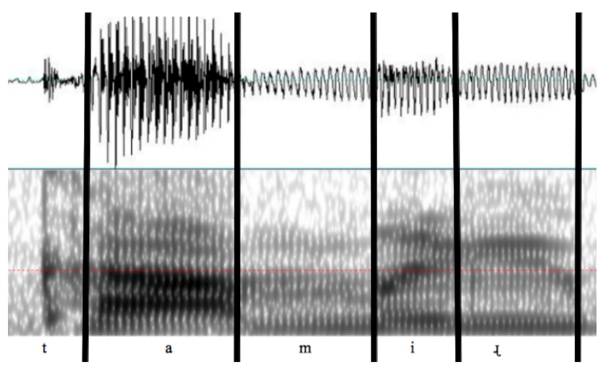

Figure 86: Waveform and spectrogram of [ɻ] in <*tamiḻ*>.

Most contemporary Tamil dialects have innovated yet a third contrastive rhotic, the postalveolar tap or trill ɾ̣ (this sound is typically transliterated as r̲; Christdas 1988:131, Narayanan et al. 1999), making for five contrastive liquids in total: ɾ, ɾ̣, ɻ, l, ɭ.[64] But ɾ̣ is clearly derived from an alveolar stop in earlier (including Middle) Tamil

[64]The posterior approximants [ɭ] and [ɻ] are often merged in contemporary dialects.

and is still pronounced as such in some conservative dialects as well as in Malayalam. Thus, for the present purposes, I consider only traditional Tamil with its two-rhotic system, as is appropriate for the corpus studies and conservative speakers considered here, and leave it open how dialects with a third rhotic might align with the present findings.

Phonotactically, the rhotics pattern as a highly sonorous natural class. For instance, only a vowel, glide, or rhotic — but not a lateral or any other consonant — can precede a geminate or cluster. This constraint is a live factor in allomorphy. For example, the dative suffix surfaces as *-kʉ* after a stem ending in an obstruent, nasal, or lateral, and as geminated *-kkʉ* elsewhere, including after vowels, glides, and rhotics, regardless of the weight of the stem-final syllable. The plural suffix *-(k)kaɭ* exhibits similar allomorphy.

Second, only nasals, laterals, and obstruents trigger progressive place assimilation (e.g. /t̪aːn/ in *kaɳ ʈaːn* vs. *t̪amiɻ t̪aːn*). Third, nasals and laterals often alternate with homorganic stops in premodern Tamil sandhi, whereas glides and rhotics never undergo such alternations. For example, laterals typically become obstruents in pre-obstruent position within the word, assimilating in (non)sonorancy, e.g. *kal* + *-pu* → *katpu* 'chastity' (cf. *caːr* + *-pu* → *caːrpu* 'place'). Fourth, poetic rhyme provides some evidence for sonority. In Tamil half-rhyme, the span of melodic correspondence normally begins with the first postvocalic consonant (Rajam 1992, Ryan 2007). But poets (especially in looser rhyme) sometimes skip over the first postvocalic consonant in assessing rhyme (e.g. *oːjn̪t̪a* ∼ *eːn̪t̪ʉ* in Kamban̠ §6,852). As in the example, skipping is most likely if the coda is a glide, the most vowel-like of the consonants. But, as Rajam (1992:193) observes, it is next most likely with the rhotics ɾ and ɻ, again suggesting that they are more vowel-like than most consonants (but less so than glides). These diagnostics are summarized in figure 87.

	glide	rhotic	lateral	nasal	obstruent
precedes geminate/cluster	yes		no		
triggers assimilation	no		yes		
alternates with stop	no		yes		
skippable in rhyme	most frequent	next most	very infrequent		

Figure 87: The Tamil rhotics as highly sonorous phonologically.

Finally, perhaps the clearest support for the consonancy and sonorancy of the rhotics comes from their phonetic characteristics, including the fact that both are spontaneously voiced and clearly liquids, being a tap and an approximant, respectively (Narayanan et al. 1999), and arguably reconstructed as such (Krishnamurti 2003). In both cases, the rhotic is not (exclusively) a coloration of the vowel (cf. the rhoticized vowels of the Dravidian language Badaga, Emeneau 1939:43ff, Ladefoged and Maddieson 1996:313ff), but a distinct constriction following the vowel, as can be seen by the formant transitions (or loss) going into the word-final segments in figures 85 and 86.

8.4 The nongeminability of rhotics

Another phonological peculiarity of the Tamil rhotics is that they are the only consonants in the language that cannot be geminated. All other consonants, including the glides and laterals, are routinely encountered as geminates and actively susceptible to gemination by phonological rules (Nagarajan 1995, Ryan forthcoming).[65] Indeed, since length is also contrastive for all Tamil vowels, it can be said that the two rhotics

[65]One other exception is the letter/phoneme called *āytam*, which is nongeminable because it cannot be an onset; but even *āytam* can undergo overlengthening (next paragraph).

are the only segments in the language that do not admit a length distinction.[66]

Furthermore, the rhotics are also the only sonorants that cannot undergo onomatopoetic overlengthening in Tamil (indicated in the script by multiplying the character, e.g. *ṯaṉṉṉena* 'pleasant, cool' in *Malaipaṭukaṭām* §352). This process, known as *aḷapeṭai*, is employed for onomatopoeia, emphasis, vocatives, metrical exigency, and so forth (Thinnappan 1976, Rajam 1992:240ff).

If a moraic representation of geminates is assumed (e.g. Hyman 1985), the weightlessness of rhotics is at least consistent with their nongeminability. Nevertheless, geminates are not always treated as heavy, and it is also the case that languages with clearly moraic rhotics sometimes prohibit specifically rhotic geminates. Sanskrit and Prakrit, for instance, permit all consonants (that can be both codas and onsets) to be geminate except for the rhotic *r*, which is repaired in sandhi (e.g. Sanskrit /punaɽ ɽaːmas/ → [punaː ɽaːmah]; Whitney 1889:§179).[67] Nevertheless, in these languages, a coda rhotic clearly confers weight to a syllable (e.g. $C_0\breve{V}R$ uncontroversially scans as heavy). Thus, the nongeminability of rhotics in Tamil, while consistent with their being weightless (or at least lighter than other coda consonants), does not provide additional independent support for that conclusion.

---

[66]Although a tap per se is unexpected on phonetic grounds to admit a length distinction, many languages exhibit a phonological length contrast in a rhotic, where the singleton reflex is a tap and the geminate a trill; furthermore, even in a language without a length contrast, a cluster of taps might still be possible, realized as a trill (Bradley 2001). Since neither is the case in Tamil, Tamil's phonological treatment of its rhotics cannot be entirely written off to their phonetics.

[67]Biblical Hebrew and Wolof are also reported to permit geminate obstruents, laterals, and glides, but not rhotics (Podesva 2002). In Hindi, the retroflex rhotics are among only a handful of segments unable to undergo gemination (Ohala 1983). In West Germanic gemination (e.g. Gothic *saljan* vs. Old English *sellan*), the rhotic is the only consonant not subject to gemination (though Old English acquired geminate *rr* through assimilations such as **ster-la* > *steorra* 'star') (Donka Minkova, p.c.).

8.5 Rhotic realization and weight

I begin with discussion of Tamil tap, as it is by far the more frequent of the two rhotics (see figure 84 for counts); I return to [ɻ] at the end of this section. Unlike many languages (e.g. Spanish) in which a rhotic exhibits markedly different allophones, e.g. being realized as a tap intervocalically and as a trill, approximant, or fricative syllable-finally, Tamil taps are traditionally realized as simple taps in all contexts, including utterance-finally. Spectrograms of two CVːR words, namely, [jaːɾ] 'who' and [saːɾ] 'sir', are given in figure 88.

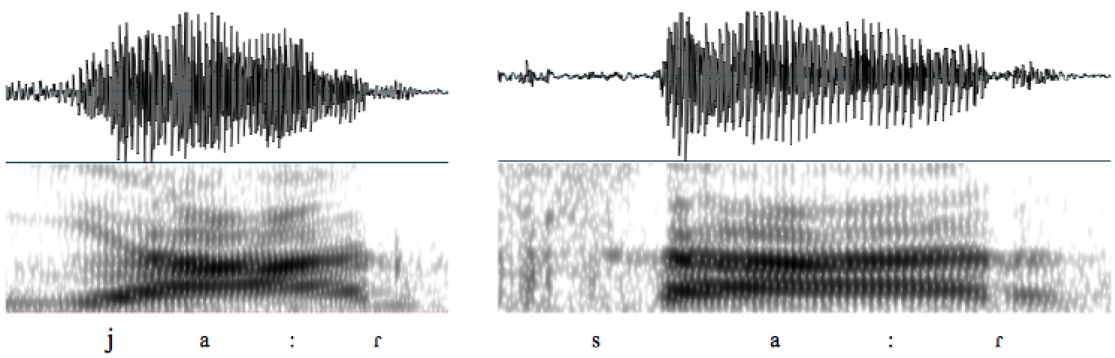

Figure 88: Waveforms/spectrograms illustrating final [ɾ] in two monosyllables.

Comparing word-final [ɾ] in monosyllables to a representative other sonorant, [n], the rhotic is significantly shorter both in raw and relative (proportional to the rime) duration, as figure 89 illustrates, based on six tokens of each word type at the end of a phonological phrase (unpaired t-test $p < .01$). As before, tap is measured from the onset of closure to the end of the release, including any svarabhakti vocalism. Moreover, the nasal is significantly longer in $C_0\breve{V}N$ than in C_0VːN (a common timing trade-off; cf. Swedish $\breve{V}$Cː vs. VːC). Thus, two factors seem to jointly contribute to the relative lightness of the tap. First, and most obviously, it is very short, so its contribution to weight is small (Gordon 2006, Lunden 2006). Second, the tap cannot be compensatorily stretched like other sonorant codas in $C_0\breve{V}C$ (recall also §8.4 on the inability of rhotics to undergo onamatopoetic extralengthening).

	mean coda duration	mean rime duration	mean coda:rime ratio
#C_0Vːɾ#	32 ms	356 ms	.09
#$C_0\breve{V}$ɾ#	N/A	N/A	N/A
#C_0Vːn#	88 ms	244 ms	.36
#$C_0\breve{V}$n#	140 ms	246 ms	.57

Figure 89: The timing of rhotics vs. nasals as codas.

The phonetic motivation for the relative lightness of [ɻ], for its part, remains an open issue (though there is no doubt that it patterns like [ɾ] in both minimality and metrics). First, it is only approximately 10% as common as [ɾ] in coda position, as mentioned above. Second, there is more dialectal variation in the realization of [ɻ]. Perhaps most contemporary speakers merge [ɻ] and [ɭ]. Third, and relatedly, descriptions disagree on the identity of [ɻ]. Hart (1999), for instance, calls it a 'lateral flap' (perhaps in consideration of the lateral/rhotic merger), and indeed, it might have been closer to a tap [ɽ] historically. Finally, being an approximant, at least in contemporary Tamil, it is more difficult to separate from the vowel for phonetic measurement than is the tap. I therefore leave the motivation for the lightness of [ɻ] an open issue. It is also logically possible that, even without obvious phonetic motivation, [ɻ] might be treated as light due to phonological symmetry among the rhotics (cf. Hayes 1997, Gordon 2002). Because the tap is considerably more frequent than [ɻ], the tap might be expected to have more gravity in phonologization.

In conclusion, Tamil $C_0\breve{V}R$ is lighter than all other $C_0\breve{V}C$, regardless of the sonority of the coda, as revealed independently by prosodic minimality and poetic metrics (and further supported by (non)geminability). In the metrics, $C_0\breve{V}R$ pattern as intermediate between light and heavy, though closer to light. Prosodic minimality, for its part, diagnoses $C_0\breve{V}R$ as categorically light.

At the same time, the rhotics are confirmed by both phonetic and phonological criteria to be both consonantal and highly sonorous, being intermediate, as typologically expected, between the glides and the laterals in sonority. Segmental weight cutoffs, it follows, are not required to coincide with sonority cutoffs. Other phonetic factors, such as the intrinsic durations of segments, can be a confound. Thus, while $C_0\breve{V}T$ patterns as lighter than $C_0\breve{V}R$ in languages such as Lithuanian (Zec 1995) and Ancient Greek (Steriade 1982; §3.2) and the reverse is found in Tamil, this dimension of crosslinguistic variation might well be phonetically grounded, given the widely varying realizations of coda rhotics.

In sum, this section focused on the perhaps unexpected lightness of $\breve{V}R$ rimes in Tamil meter. I first demonstrated that this special treatment of $\breve{V}R$ as lighter than other $\breve{V}C_{\neq R}$ is also diagnosed by at least one other phonological system, namely, prosodic minimality. I then explained this treatment of rhotics as (relatively) light in terms of phonetic duration, in particular, the fact that rhotic codas in Tamil, unlike in certain other languages, tend to be shorter than other codas, rendering the whole $\breve{V}R$ rime comparatively short.

9 On the general interface of phonetics and metrical weight

9.1 Tamil metrical weight vs. rimal duration

A preliminary hypothesis is that (gradient) metrical weight is driven by the phonetic duration of the rime. In this section, I show that this hypothesis is in general highly accurate, though it incorrectly predicts that laterals should be lighter than nasals in Tamil. In other words, rime duration is a good predictor of metrical weight but unlikely to be the only phonetic factor (as followed up in the next section, §9.2).

As a preliminary investigation of the relation between gradient weight in metrics and the phonetic characteristics of syllables, I collected phonetic information on 351 consecutive syllables in a recording of high-register, conservative Tamil (specifically, Kausalya Hart reading passages from her Tamil textbook, Hart 1999). Though several centuries separate this recording from the composition of Kambaṉ's epic, which I continue to employ as a metrical corpus, Tamil pronunciation, particularly in its highest register (known as *cen-tamiḻ*), has evidently changed relatively little over this span, and serves as a reasonable approximation for exploratory purposes.[68] To the extent that pronunciation has changed, this comparison is overconservative: One would expect the correlations reported below to be (if anything) stronger if the phonetic and poetic corpora were more closely aligned.

As an initial assessment of the relation between phonetics and metrical weight, I plot in figure 90 the mean durations of rimes in ms (x-axis) against their mean weight propensities in the metrics (y-axis) (on the connection between duration and syllable weight, see Maddieson 1993, Hubbard 1994, Broselow et al. 1997, Gordon 2006). The weight of each rime type was estimated from the metrical corpus using the linear regression model in §2.5.4 (also discussed in §7), except with individual

[68]Evidence for the conservatism of this acrolect, from which the local dialects have diverged considerably (Tamil is diglossic), comes from several quarters. For one, the modern orthography was largely settled by Kambaṉ's time, and the pronunciation remains very close to that orthography, despite grammatical changes. Departures do occur, however; for instance, Ms. Hart often harmonizes short medial vowels in a way that is not orthographically indicated (but such an innovation has little bearing on weight). Furthermore, phonetic treatises on the language exist from Kambaṉ's period and earlier (e.g. *Naṉṉūl*; bibliography in Rajam 1992), and these accounts are broadly consistent with present-day orthoepy (e.g. the special status of [ăj] is acknowledged).

rime types assessed as factors. The range on this axis is close to $(0, 1)$ because the model's log-odds estimates are translated into probabilities. Duration was measured following Gordon (1999, 2002) using discontinuities in the spectrogram and waveform. For syllables closed by a geminate, the midpoint of the geminate was taken to be the endpoint of the rime. Some horizontal stratification is visible due to repeated rime types in the phonetic data, which all correspond to the same metrical propensity. These data do not take word shape or position in the word into account, though one might expect the the distributions to be similar in the two corpora, given that they are (loosely speaking) the same language.

Even with these compromises (no consideration of perceptual energy or contextual effects of word position, arbitrarily bifurcated geminates, and ignoring dialectal/diachronic discrepancies), a high correlation between phonetics and metrics is already observed, as indicated by the longest (red) regression line in figure 90 (Pearson's product-moment correlation $r = .840$, Spearman's rank correlation $\rho = .847$, both $p < .0001$). Moreover, these correlations are not driven exclusively by the heavy/light difference. Even within the sets of heavies and lights considered independently, significant positive correlations hold, as indicated by two shorter (blue and green) regression lines (r and $\rho > .45$ for both, all $p < .0001$).

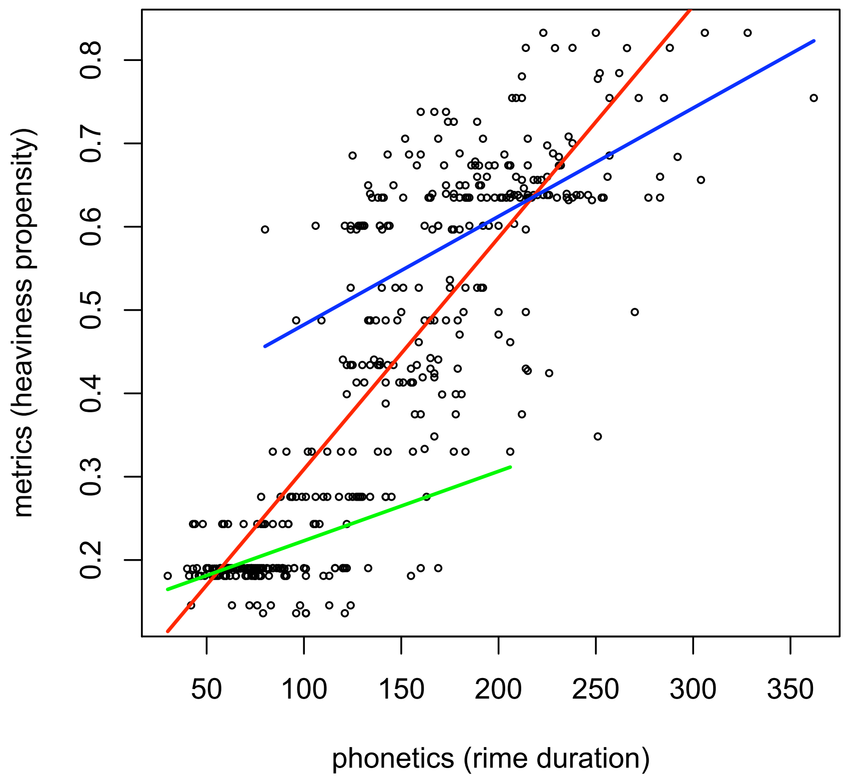

Figure 90: Metrical weight as a function of rime duration.

Figure 91 shows rime types instead of tokens as in figure 90. Only types with five or more tokens in the phonetic data are shown ($N = 23$). The symbols 'L' and '1' represent [ɭ] and [ʉ], respectively. With this greater degree of abstraction, the correlations tighten (overall correlation $r = .918$, $\rho = .915$, both $p < .0001$).

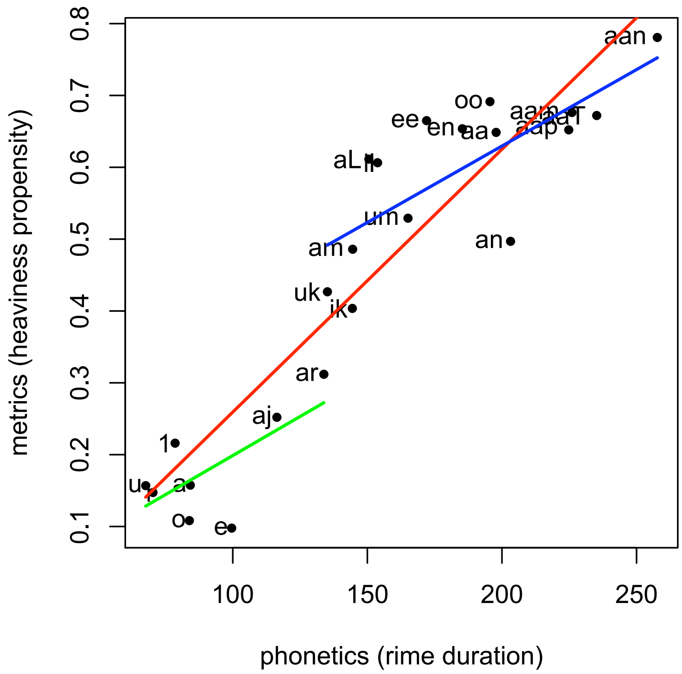

Figure 91: Metrical weight vs. duration (types).

Finally, figure 92 collapses the rimes in figure 90 (returning to the full data set) into broader phonological classes, using the same categories as in §7 (though two of the categories, V̆W and VVCC, are left out because they are absent from the phonetic data). The values on both axes are weighted averages of the values in figure 90. For example, the rime *aɭ* is three times as frequent as *il* in the phonetic data, therefore the former is given three times as much weight in determining both the x and y positions of 'lateral' (i.e. V̆L). The correlation continues to tighten (r = .930, ρ = .976, both p < .0001), and it is now clear that the category hierarchies are in approximate agreement between the phonetic and metrical diagnostics. For example, 'diphthong' (i.e. [ăj]) and 'rhotic' (i.e. V̆R) are the two categories nearest to V̆, respectively, on both dimensions. Thus, the subhierarchy V̆ < ăj < V̆R < V̆T is significant not only in the metrics (§7), but also in the phonetics (respective sample sizes of the four categories: 171, 28, 19, 63; respective t-test one-tailed p-values for the three contrasts: $p < .001$, $p = .020$, $p = .018$). However, V̆L patterns as heavier than duration predicts.

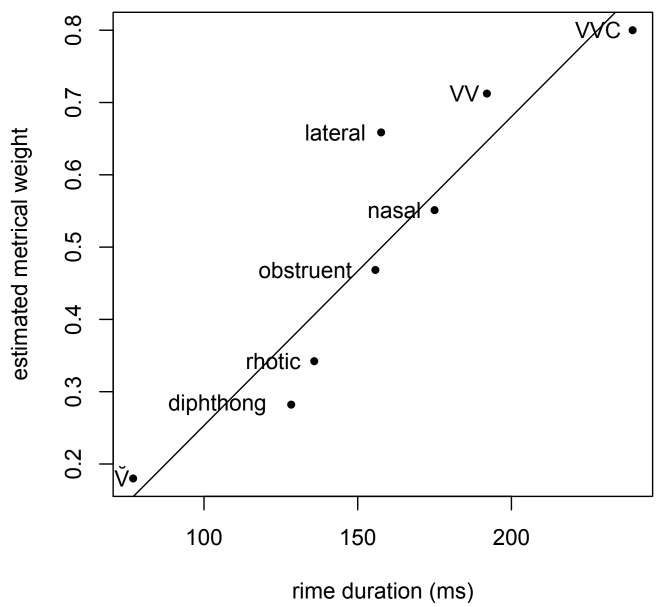

Figure 92: Metrical weight vs. duration (broader category means).

This discrepant ordering of the nasals and laterals is found independently across positions of the word, as figure 93 reveals, which splits figure 92 into three plots, one for each of initial, medial, and final position (all excluding monosyllables). Two types are missing in initial position: ‘diphthong’ (i.e. [ăj]), which does not occur in initial position (its long counterpart [aj] does, but [aj] is absent from the phonetic data), and ‘rhotic’ (i.e. V̆R), an accidental gap in the present phonetic data.

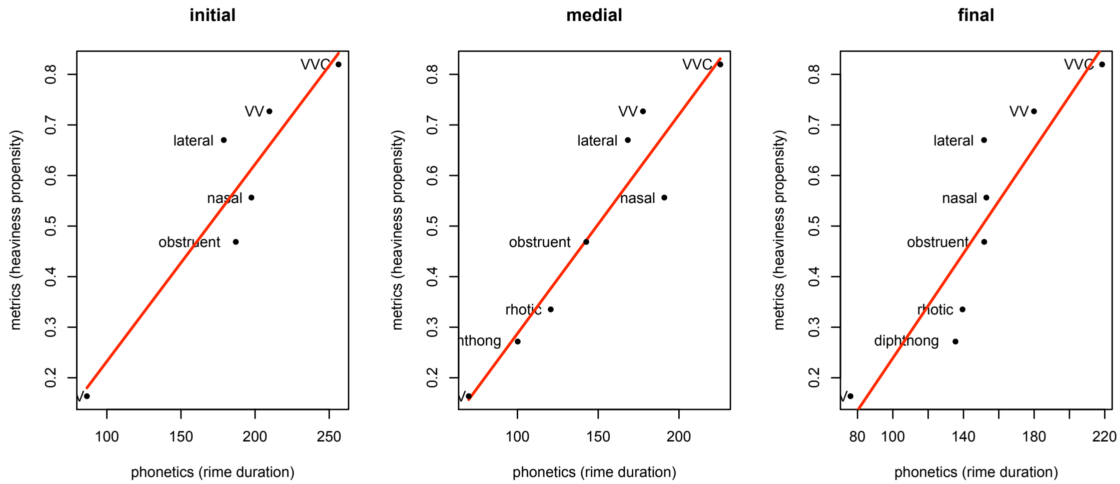

Figure 93: Metrical weight vs. duration in initial, medial, and final positions.

9.2 A role for intervals and/or energy?

In §9.1, despite achieving overall correlations of $r > 0.9$ (e.g. $r = .93$ for eight categories in figure 92), some discrepancies between rimal duration and inferred metrical weight suggest that the rimal duration alone might not be the optimal phonetic model. In particular, V̆N rimes were significantly longer than V̆L rimes, even though the latter pattern as significantly heavier in the metrics, a discrepancy observed independently in word-initial, -medial, and -final syllables. In this section, I discuss two possible directions for improving the phonetic model, improving the overall metrics-phonetics fit and eliminating the V̆L-V̆N mismatch. First, I consider the duration of a span more inclusive than the rime, namely, that of the vowel-to-vowel interval. Second, I consider incorporating perceptual energy into the model in addition to duration.

The goal of the present discussion is not to reach a decisive answer concerning the phonetic interface of (gradient) weight, but rather to describe two of the main issues for the future development of such a model, namely, the span over which weight is computed (e.g. rime vs. interval, or the precise nature of syllabification) and the proper treatment of the integration of perceptual energy and duration (including, most generally, the extent to which energy exerts an independent effect; cf. Gordon 2002, 2006, Gordon et al. 2008). While the work just cited focuses on optimally partitioning phonetic data into weight categories, evaluating gradient weight systems, as I am doing here, opens up a new empirical field in which the correlation between the phonetic model and the weight-sensitive phonological system can be directly investigated as such.

First, I consider the span over which duration is assessed, which was assumed to be the syllable rime in §9.1. Steriade (2008b, 2009, 2011; cf. also 2008a) has proposed that quantity in verse is based not on syllabic intervals, but on total vowel-to-vowel intervals (or simply INTERVALS), which span from the beginning of each vowel to that of the next (or to pause at the end of the phonological domain). Figure 94 shows syllable (top) vs. interval (bottom) analyses of the word *structure*, uttered in isolation by an American English speaker (corresponding to the underlying form /stɹʌkʃɚ/ as opposed to /stɹʌktʃɚ/, another variant). On the interval analysis, [stɹ] is extraprosodic, while [ʃ] is affiliated wholly with the first interval. An onset such as [ʃ] is expected to play a role in quantity on this proposal — but only of the interval headed by the preceding nucleus. An onset cannot, on this proposal, contribute to the quantity of the metrical constituent headed by the following nucleus. Steriade adduces a range of evidence for this proposal, including purported explanations of final extrametricality, timing compensation, rhyme span structure, and so forth (*ibid.*). See also §6 for references on intervals in Norse metrics.

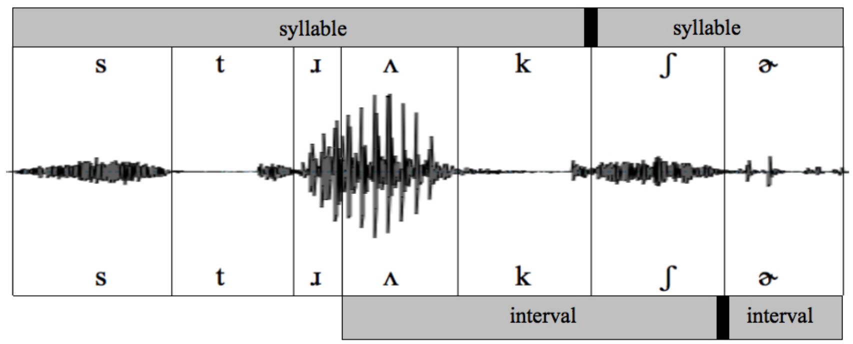

Figure 94: Syllables vs. V-to-V intervals for <*structure*>.

As another illustration of syllables vs. intervals, the first couplet of Kamban̲'s epic (repeated from figure 2) is parsed into syllables (top) and intervals (bottom) in figure 95. As usual, periods indicate syllable boundaries; I employ bars for interval boundaries. Corresponding weight templates are given to the right. The binary criterion for syllables is the traditional one: In terms of timing slots of the rime, X is light, X_2 is heavy (subscripts indicate lower bounds, superscripts upper bounds). The corresponding binary criterion for intervals could be X^2 is light, X_3 is heavy. The weight templates are then the same on both analyses, with the exception of the final position, which is light in terms of intervals but heavy in terms of syllables. Nothing in these examples favors one analysis over the other.

(a) syllable parse

u.la.kam .jaː.ʋăj.jun̪ .t̪aː.m u.ɭa .ʋaːk.ka.lum	LLHHLH	HLLHLH
n̪i.lăj .pe.tut̪.t̪a.lʉ .n̪iːk.ka.lʉ .n̪iːŋ.ka.laː	LLLHLL	HLLHLH

(b) interval parse

ul|ak|am j|aːʋ|ăjj|un̪ t̪|aːm |uɭ|a ʋ|aːkk|al|um	LLHHLH	HLLHLL
n̪|il|ăj p|et|ut̪t̪|al|ʉ n̪|iːkk|al|ʉ n̪|iːŋk|al|aː	LLLHLL	HLLHLL

Figure 95: Syllables vs. intervals for a Tamil couplet.

In the Tamil phonetic data (§9.1), the durations of V̌NC intervals are approximately the same as those of V̆LC intervals, despite V̆N rimes being longer than V̆L rimes. Token boxplots for the durations of V̆N and V̆L as rimes (left) vs. intervals (right) are given in figure 96. Only non-word-final syllables/intervals are considered here.[69] Thus, the durations of V̆N and V̆L are more strongly correlated with their weights under intervals than under syllables.

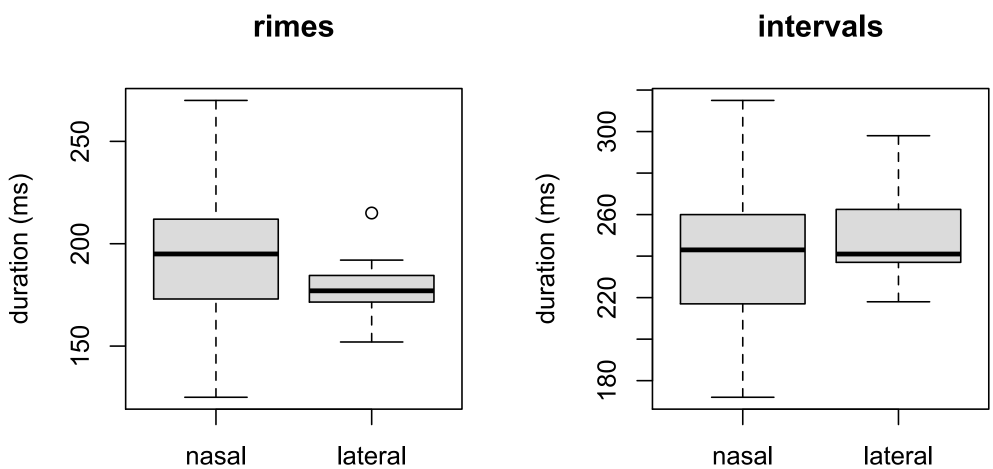

Figure 96: V̆N vs. V̆L: rimes (left) vs. intervals (right).

On inspecting the data, the vast majority of V̆N rimes are in homorganic nasal-stop clusters, while the vast majority of V̆L rimes are in geminates (most commonly from the lemma *illăj* 'no(t)' and related forms). The same generalization holds for the metrical corpus. Recall that in measuring the durations of codas, geminates were bifurcated at the midpoint, so that most lateral codas would be measured as

[69]This exclusion is due to a methodological ambiguity concerning measuring word-final intervals. In particular, it is sometimes unclear whether to count a particular word boundary as a pause, in which case the interval is identical to the rime, or not, in which case the interval extends over the boundary to the onset of the following vowel. I avoid this question altogether by employing only non-word-final intervals. For fairness of comparison, I also exclude word-final rimes in the syllable set, though they are unambiguous in this respect.

approximately 50% of their (geminate) interludes. In nasal-stop interludes, on the other hand, the nasal coda occupies on average 82% of the interlude, as in the example in figure 97. Under interval theory, there is no question of how geminates should be treated by the phonetic model: The whole geminate is included in the interval with the preceding vowel.

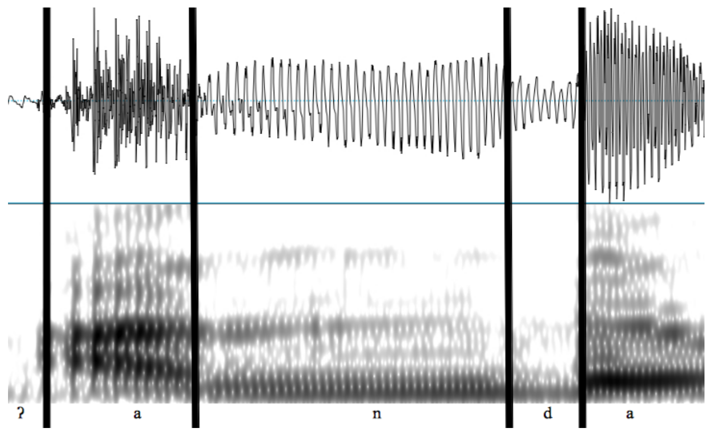

Figure 97: Waveform/spectrogram for Tamil <*anta*> [ʔan̪d̪a].

Figure 98 is organized just like figure 92, except now using interval duration instead of rime duration. The correlation improves for intervals, in part because V̆N and V̆L assume the same order by both diagnostics ($r = .971$, $\rho = 1$, both $p < .0001$).

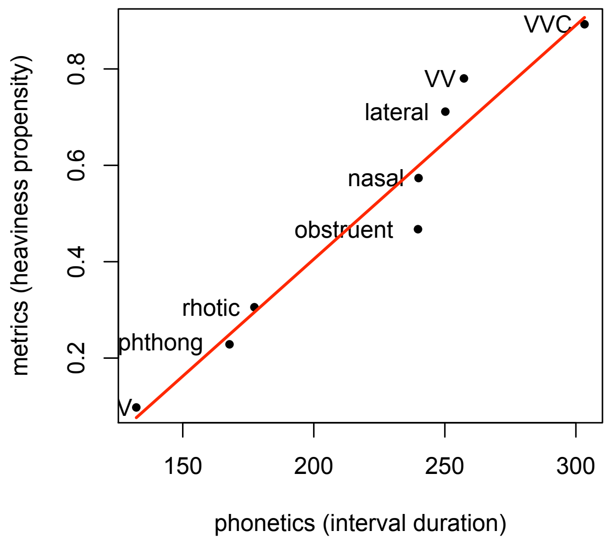

Figure 98: Metrical weight vs. duration with intervals.

Another (not mutually exclusive) potential explanation for the treatment of rimal V̆L as heavier than V̆N even though the latter is longer in duration is phonetic energy, probably the most consistent non-durational phonetic correlate of sonority (Parker 2002). Laterals are typically more energetic than nasals, as the data below will corroborate for Tamil. Gordon (2002, 2005, 2006) argues that perceptual energy (integrated over duration) plays a role in grounding categorical weight distinctions. Moreover, if a distinction is made, the rime V̆C (at least if C is supralaryngeal, cf. Crosswhite 2006) is virtually always lighter than VV, even when V̆C is longer, on average, than VV (cf. Ohsiek 1978, Hoequist 1985, Leskinen and Lehtonen 1985, Devine and Stephens 1994:72), presumably because C slots are less energetic than V slots.

Figure 99 contains two bar charts. The one on the left shows the average mean-energy intensity reduction in dB SPL (decibels sound pressure level) as measured in

Praat (Boersma and Weenink 2011; see manual for formula) from the vowel into each of four types of codas, drawing on the Tamil phonetic corpus (§9.1). For example, a typical obstruent coda has a mean-energy intensity (hereafter, simply ENERGY) of 10.5 dB less than the preceding vowel. As this chart reveals, laterals have the smallest drop-off (i.e. are the most vowel-like in energy), followed by the nasals, rhotics, and obstruents, respectively.[70] Putting aside the rhotics (fn. 70), we thus see the sonority hierarchy: obstruent < nasal < lateral < vowel.

The dB reductions in the left plot in figure 99 can then be approximately transformed into proportional reductions in (perceived) loudness, as illustrated in the right plot. The transformation employed here is the same one used by Gordon (2002), which draws on the experimental results of Warren (1970) (see also Moore 1989); note that this is only a rough approximation of loudness, though sufficient for the present purposes (see Gordon 2005, Gordon et al. 2008 for caveats and refinements). For example, the 10.5 dB reduction for obstruents corresponds to a 56% reduction in perceived loudness.

[70] The energies of rhotic tokens are less consistent than those of the other categories, given that heterogeneous events, including the tap itself and the optional svarabhakti release, are measured together as the segment here. Energy is thus sensitive to the tap:svarabhakti ratio, as well as different degrees of closure for the tap, resulting in more variance for rhotics than other types.

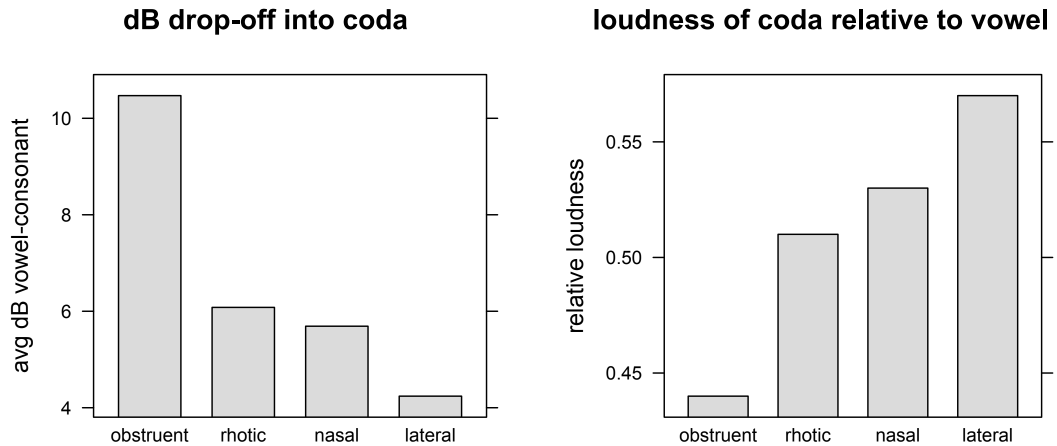

Figure 99: Energy/loudness of coda relative to preceding vowel.

For each rime or interval token, I compute a loudness-adjusted weight percept according to the equation in figure 100 (cf. total perceptual energy in Gordon 2002:§7.3). For each rime/interval token, the duration of the vowel is summed with the loudness-adjusted duration of the coda, if any (or whole interlude, in the case of intervals). The energy of the vowel is used as a baseline for normalization, controlling for variation in speech volume across tokens.[71] The estimated loudness of the coda relative to the preceding vowel is then computed from the relative energies of the respective segments (as in figure 99). As codas are normally less energetic than vowels, this estimate is typically a fraction less than one (e.g. an average of 0.44 for obstruents). Finally, a constant (c in figure 100) modulates how much weight energy is given relative to duration in determining the final percept. Because c is a denominator, as it approaches infinity, energy approaches having no effect on weight. As c approaches zero, on the other hand, the relative loudness of the coda plays an increasingly dominant role.

[71]Because vowels are used for normalization, this method irons out any intrinsic differences in energy between vowels. Nevertheless, the focus in this section is coda-driven differences.

Let	V_0ms	be the duration of the vowel (ms),
	V_0dB	be the mean-energy intensity of the vowel (dB),
	C_0ms	be the duration of the coda (ms),
	C_0dB	be the mean-energy intensity of the coda (dB),
	Δloudness$(x-y)$	be a function from Δenergy$(x-y)$ to estimated Δloudness,
	c	be a constant, $0 < c < \infty$.

$$\boxed{\begin{array}{l}\text{loudness-adjusted weight} = \\[1em] \mathrm{V_0ms} + \mathrm{C_0ms} \cdot (1 - \frac{1 - \Delta\text{loudness}(\mathrm{C_0dB} - \mathrm{V_0dB})}{c})\end{array}}$$

Figure 100: Equation for loudness-corrected duration of a rime token.

For example, consider a token of the rime [ik], in which the energy and duration of [i] are 60.3 dB and 50 ms, respectively, and the energy and duration of [k] are 52.4 dB and 81 ms. The consonant is 7.9 dB less than the vowel, or ~0.48 times as loud. Let us say c is 2, so that the contribution of coda energy is attenuated somewhat with respect to that of duration. The coda duration (81 ms) is then multiplied by the adjusted loudness $(1-(1-.48)/2 = .74)$, giving 59.9. Finally, the vowel duration is added, giving a loudness-adjusted duration of the rime of 109.9 in arbitrary units (instead of the 131 ms obtained from raw duration).

Factoring in loudness in this manner can only slightly improve the metrical-phonetic correlation for rimes. The value of c maximizing this correlation (for the eight rime categories in figure 92) is 5.2, for which $r = .933$, only a slight improvement over $r = .930$ for the pure-duration rime model in figure 92. In this model, $\breve{\mathrm{V}}$L still significantly trails $\breve{\mathrm{V}}$N in the phonetics, contrary to the metrics. If the contribution of energy is increased (by decreasing c) to the point where $\breve{\mathrm{V}}$L overcomes $\breve{\mathrm{V}}$N in the phonetics, the overall correlation is substantially worse (specifically, $\breve{\mathrm{V}}$L crosses $\breve{\mathrm{V}}$N at $c = .45$, at which point $r = .21$). In other words, while $\breve{\mathrm{V}}$L is more energetic than $\breve{\mathrm{V}}$N, any rime model that weights energy highly enough relative to duration in order for V̆L to be treated as heavier than V̆N will fare substantially worse in general, creating more problems than it fixes.

Intervals can be improved more substantially by taking energy into account, adjusting the equation in figure 100 to include both the coda (if any) and the immediately following onset (if any). Recall that the pure duration interval model in figure 98 had a correlation of $r = .971$ ($\rho = 1$). Using the energy-adjustment formula with the optimal c (2.64), the error is almost cut in half ($r = .984$, $\rho = 1$), though either way, the correlation is almost at ceiling.[72]

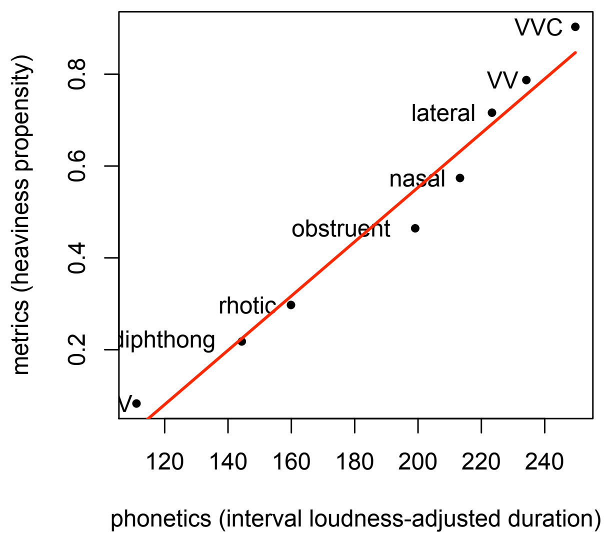

Figure 101: Metrical weight vs. duration with energy-adjusted intervals.

In conclusion, the optimal phonetic model tested here for Tamil, in the sense of best correlating with the weights inferred from the metrical corpus (ignoring the

[72]The correlation for tokens, as in figure 90, as opposed to categories, is also improved, at $r = .857$, though this coefficient cannot be compared directly to the $r = .840$ coefficient in §9.1, as the intervals are drawn only from non-final positions and the metrical propensities for intervals are based on fewer data than those for rimes, which are less specific.

chronological difference), is one based on vowel-to-vowel interval duration with a small adjustment for loudness, such that less energetic consonants contribute less to weight than more energetic ones. In the case of the meter(s) of Kamban̲'s epic, at least, the correlation between phonetics and metrics appears to be rather tight, given that even duration alone was giving correlations of $r > .9$, which refinements improved to $r > .98$ (for categories). In the following section, I consider a rather different situation in Finnish.

9.3 On the phonetic interface in Finnish

As a representative sample of Finnish pronunciation, I measured 162 word-initial (primary stressed) syllables harvested from a website featuring WAV files of a native Finnish speaker (Kai Nikulainen) pronouncing Finnish words for pedagogical purposes (`www.sci.fi/~kajun/finns/Samples`, accessed May 2011). Like the Tamil phonetic data, this is a modern speaker's pronunciation, while the verse corpus is more conservative (in this case, compiled in the mid-19th century but with earlier antecedents). The idea is to provide only an approximation of the phonetics of the verse language as a starting point for the discussion of the phonetics/metrics interface; the details of the phonetic model are not crucial here. Moreover, as with the Tamil, insofar as the speaker's pronunciation departs from the Kalevala pronunciation, the comparison here is overconservative, not underconservative.

The correlation between rimal duration and metrical weight (as inferred from strong vs. weak skews of stressed syllable types as in §4.2.3) is plotted in figure 102 for tokens (left) and categories (right), using the same four skeletal categories employed in §4.2.3, where it was demonstrated that the meter observes the following hierarchy: $\breve{\mathrm{V}} < \breve{\mathrm{V}}\mathrm{C} < \mathrm{VV} < \mathrm{VVC}$ (as is also evident in the vertical alignment of categories in the plot on the right). Within the set of heavy tokens in the plot on the left, a small but significant correlation is observed (Pearson's $r = .226$, $p = .009$; Spearman's ρ positive but non-significant), as depicted by the uppermost (blue) regression line. No significant correlation obtains within the light tokens.

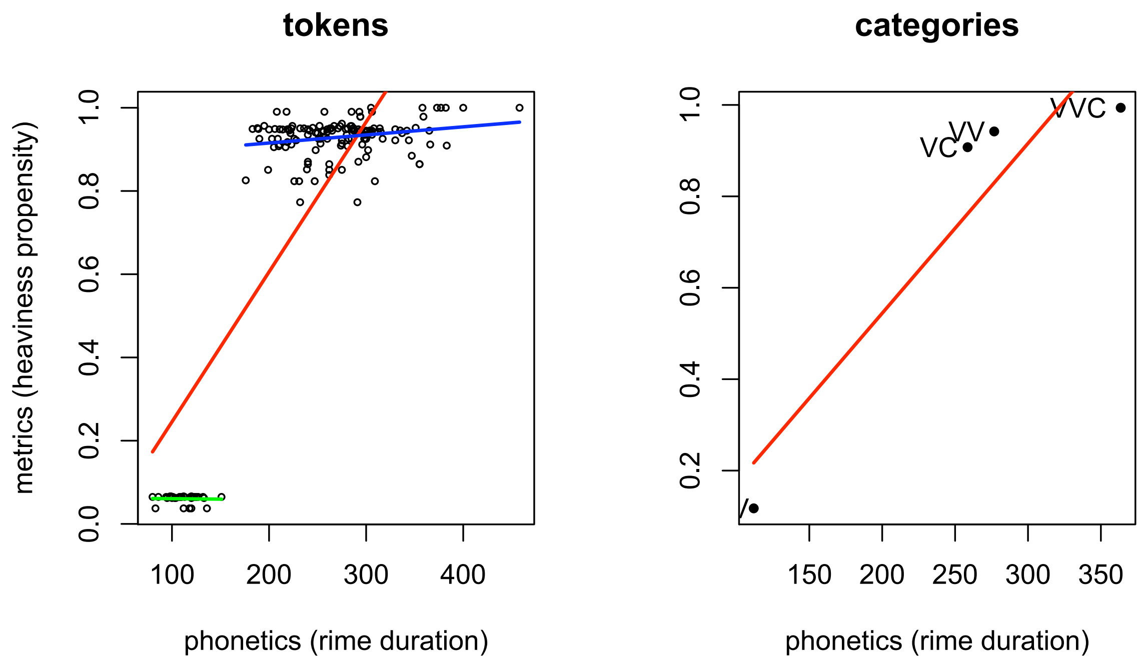

Figure 102: Metrical weight vs. rime duration in Finnish.

I now return to the optimal model tested above for Tamil, namely, one in which duration is computed over the whole vowel-to-vowel interval rather than just the rime and in which the contribution of each consonant is corrected for its estimated loudness relative to the preceding vowel (using the same c as for Tamil; see §9.2). For both tokens and categories, the correlations improve, though only slightly, as shown in figure 103. In particular, within heavy tokens, $r = .262$ (vs. $r = .226$ above); within lights the correlation remains non-significant. For categories, overall $r = .948$

(vs. $r = .932$ above).

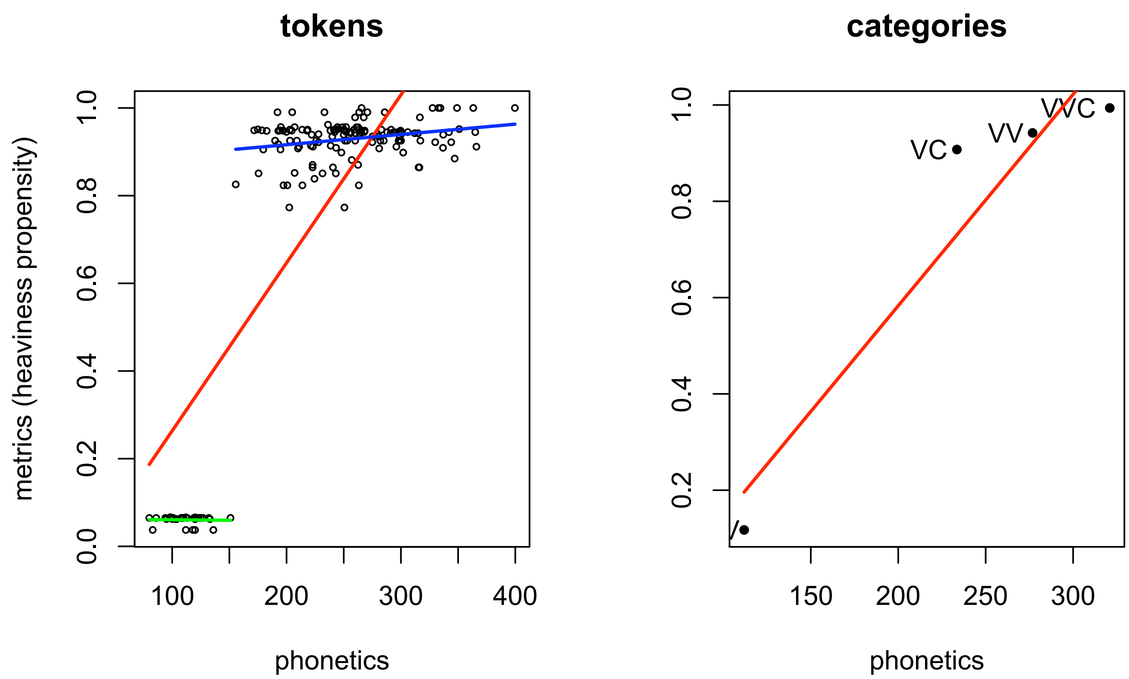

Figure 103: Metrical weight vs. phonetic weight in Finnish.

Crucially, regardless of the phonetic model (whether rimes or intervals, or whether raw duration or duration corrected for loudness), the same generalization emerges for Finnish: A much greater gap is found between light ($C_0\breve{V}$) and heavy (other) syllables in the metrics than is predicted by a phonetic model based on duration and/or loudness. This discrepancy is clearest in the token plots above, in which the dots are relatively continuous along the phonetic dimension (with heavy and light almost overlapping), while the metrics draws a sharp distinction between them. Put differently, there appears to be more categorical polarization in the metrics than the phonetic model implies. Furthermore, this polarization appears to be language- or corpus-specific, given that no comparably substantial polarization was observed for

Tamil in §9.1. I propose a generative model for this situation in §10.

10 Hybrid categorical-gradient systems: generative analysis

10.1 Varying categoricity in the treatment of weight

I now turn to the treatment of gradient weight in a generative model of weight mapping (presently for poetic meter, though the same principles apply to the analysis of gradient weight in stress in part IV). Following work in generative metrics (e.g. Halle and Keyser 1971, Halle 1970, Kiparsky 1977, Hayes 1988, Hanson and Kiparsky 1996), I assume that meters comprise abstract templates of strong and weak positions (see §4, §6), among other structure. Constraints indexed to these positions define the sets of syllables permitted in them (e.g. *heavy/W: 'penalize a heavy syllable in a weak position'). By WEIGHT MAPPING, I refer to this correspondence relation between metrical positions and syllabic weight.

I begin by schematizing a space of logical possibilities. At one extreme, weight mapping might be fully categorical. Figure 104 illustrates hypothetical binary (left) and ternary (right) categorizations of dummy data (sampled from normal distributions) on an arbitrary perceptual-phonetic scale (x-axis).[73] These plots are organized like figure 90 above, in which the y-axis is weight (a phonological variable) and the x-axis duration (a phonetic variable). The categories in figure 104 overlap somewhat

[73] Yana stress (stress the leftmost heavy, if any, otherwise initial; Sapir and Swadesh 1960) exemplifies a fully categorical binary system. Kashmiri stress (Morén 2000) is ostensibly a ternary case. If one were to plot the lengths of rimes in these languages against their categories, patterns like the ones exemplified would be expected.

on the phonetic dimension, representing, let us say, a bias for phonological simplicity in categorization at the expense of fidelity to the phonetic scale (Hayes 1997, Gordon 2002).

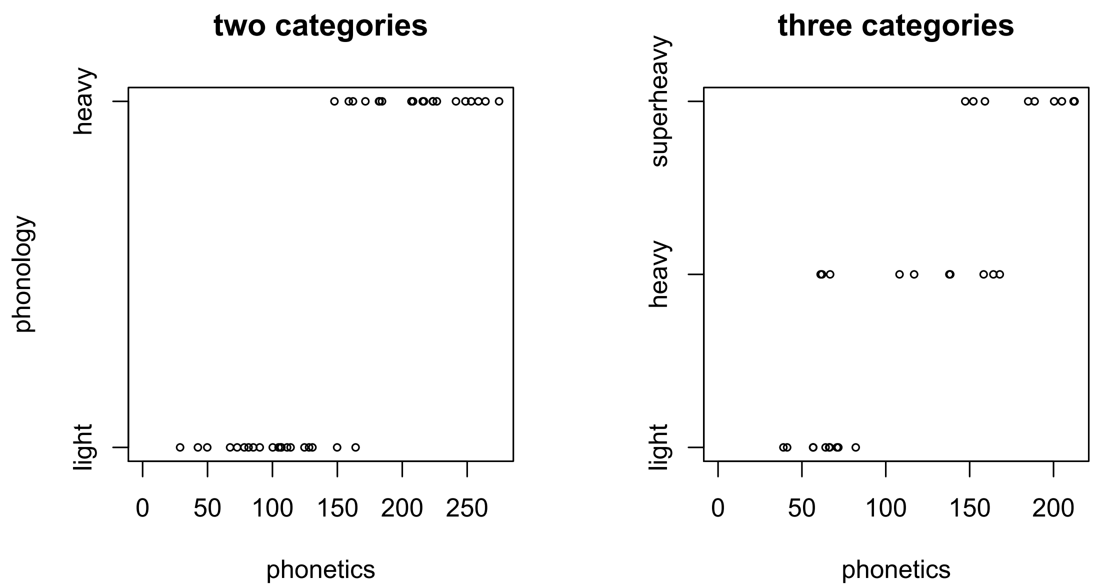

Figure 104: Phonetics-phonology interface under total categorization.

At the other extreme, one can imagine that the phonology (y-axis) might directly reflect the relevant perceptual-phonetic scale without any interference from categorization (cf. Flemming 2001, 2003, 2004, Steriade 2000, 2001a), as illustrated by the first (top left) plot of figure 105. In this plot, the correlation between phonetics (unlabeled x-axis) and phonology (unlabeled y-axis) approximates the diagonal (adding noise for realism). As the other plots in figure 105 suggest, one can also imagine situations intermediate between these two extremes, in which the phonology is polarized towards categories, but within categories continues to reflect gradient sensitivity to a phonetic scale. I term this intermediate situation INCOMPLETE CATEGORIZATION (not to be confused with overlapping but complete categorization, as in figure 104). The plots in figure 105 are arranged from the least categorical (plot 1) to the most (plot 9; as in figure 104).[74]

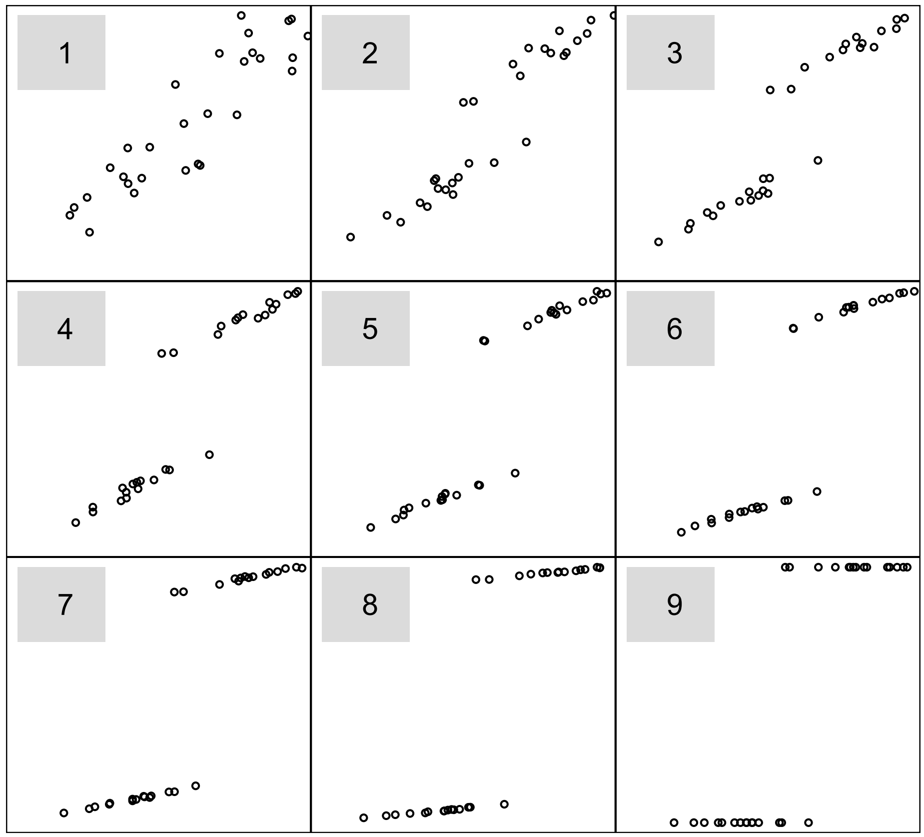

Figure 105: A continuum from no (plot 1) to total (plot 9) categorization.

Meters evidently vary in the extent to which they treat weight as binary-categorical as opposed to gradient, instantiating different points along the continuum of categoricity sketched in figure 105. Impressionistically, weight appears to be highly (though not exclusively) categorical in Homer's epics (§3.2) and the Kalevala (§4), but less so in Kamban̲'s epic (§2), in which it is not obvious that categorization plays any role. More concretely, compare the Tamil (left) and Finnish (right) distributions

[74]Similar distributions are analyzed in connection to experiments on categorical perception (e.g. DiCanio forthcoming and references therein).

in figure 106. In both plots, the means for V̆, V̆C, VV, and VVC are superimposed on gray points representing rime tokens. The Tamil data are repeated from figures 90 and 92, while the Finnish data are repeated from figure 103, both with the same organizations as previously.[75] In both plots, the data are relatively continuous on the phonetic dimension (as also seen in Gordon's 1999 Finnish data). They differ mainly on the metrical dimension: The Tamil data are continuous, with V̆C (whether including or excluding V̆R; §7.2) being approximately halfway between V̆ and VVC. The Finnish data, by contrast, are widely separated, with V̆C being over 10 times as close to VVC as it is to V̆.[76] At the same time, weight is not exclusively binary in Finnish, as the contrasts in V̆C < VV < VVC are significant (§4).

---

[75]The speech rate is generally slower in the Finnish data than in the Tamil data, accounting for the broader range on the x-axis in the Finnish.

[76]Categoricity is orthogonal to the range on the y-axis, which corresponds to the overall strictness of the meter. If the difference between strong and weak positions were attenuated, the points would approach a horizontal band. Thus, while the strictness of the meter corresponds to the slope of the regression line in such a plot, categoricity concerns only the degree of polarization on that continuum.

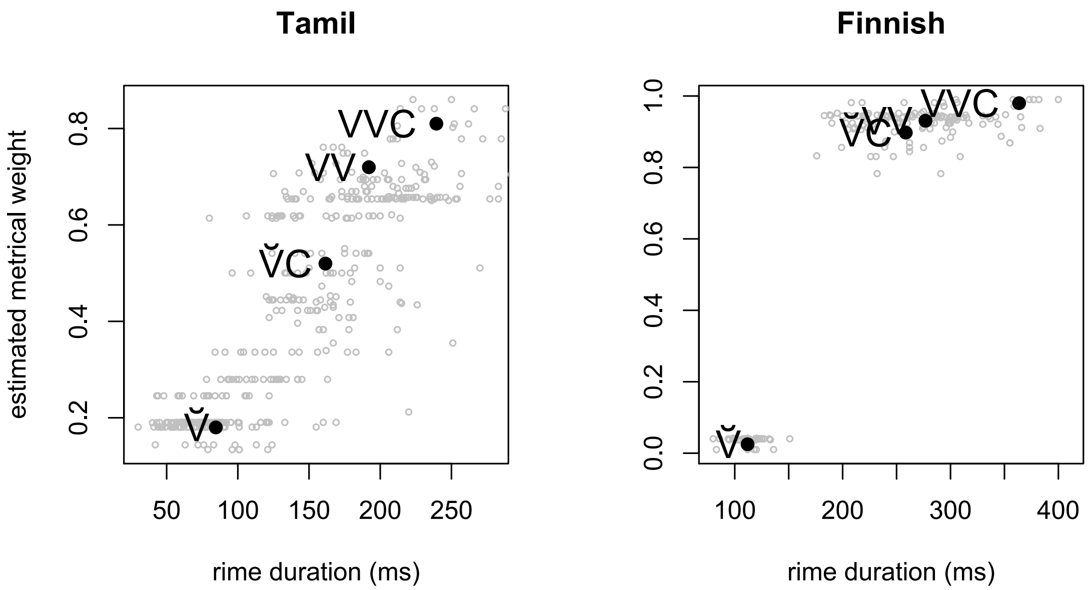

Figure 106: Differing categoricity in Tamil (Kamban̲) vs. Finnish (Kalevala) meters.

As a constraint framework capable of handling gradient variation, I employ maximum entropy (maxent) grammar (Johnson 2002, Goldwater and Johnson 2003, Wilson 2006, Jäger 2007, Hayes and Wilson 2008, Ryan 2010, Hayes and Moore-Cantwell 2011; cf. Boersma and Pater 2008), a type of Harmonic Grammar (Smolensky and Legendre 2006, Pater 2009, Potts et al. 2010), but with probabilistic interpretation of candidate harmonies and often a Bayesian approach to constraint weighting. As in all varieties of Harmonic Grammar, the score of each candidate is the sum of its weighted violations. In maxent grammar, rather than treating the candidate with the greatest score as the categorical winner, all scores are transformed into probabilities (though often ones so close to zero or one that the output is effectively categorical). Specifically, if s_i is a candidate's nonnegative penalty score, $p(\text{candidate}) = e^{-s_i} \ / \ \sum_j e^{-s_j}$, where j ranges over all candidates. Constraints are weighted so as to maximize the likelihood of the observed distribution of outputs, possibly with smoothing. See the references for a more detailed overview.

Incomplete categorization can be modeled in such a framework by the interaction of categorical and scalar constraints. A fully categorical distribution (as in plot 9 in figure 105) is modeled by giving the relevant categorical constraints positive weights and scalar constraints zero weights. A direct interface distribution (as in plot 1) can be achieved by giving scalar constraints positive weights and categorical constraints zero weights. Incomplete categorization emerges if both categorical and scalar constraints are given positive weights (e.g. plots 2 through 8, according on their relative weights).

A categorical constraint (e.g. *heavy/W: 'penalize a heavy syllable in a weak position') is either violated or not for each locus of violation; it cannot be violated multiple times (or to some other degree) by a single locus (McCarthy 2003). A scalar (also known as gradient) constraint, by contrast, is violable to a greater or lesser degree by a single locus (e.g. Flemming 2001, Pater forthcoming). In the present case, scalar constraints are labeled with the comparative suffix (e.g. *heavier/W) and violated to a real-valued degree supplied by the phonetics (as in Flemming 2001; cf. the more abstract integer scale for sonority implied by HNuc; Prince and Smolensky 1993/2004, McCarthy 2003, Pater forthcoming).[77] For the sake of illustration, the phonetic model here is based on the log mean duration of the rime, though a more accurate model would likely be perceptually grounded (see §9.2), integrating loudness, auditory recovery, and so forth (Gordon 2005, Crosswhite 2006, Gordon et al. 2008).

Figure 107 is a maxent tableau illustrating four candidate lines for the Kalevala meter. Lines (a) and (b) are actually from the Kalevala while (c) and (d) are oc-

[77] While McCarthy (2003) argues for the exclusion of scalar (in the present sense) constraints from Optimality Theory, Pater (forthcoming) maintains that they are appropriate for (serial) Harmonic Grammar. Flemming (2001) is also couched in Harmonic Grammar, not Optimality Theory.

tosyllabic snippets extracted from prose, included for comparison. Four mapping constraints are given in the top row in decreasing order of weight. For the Kalevala meter, they evaluate only stressed (i.e. word-initial) syllables, excluding those in line-initial position (cf. Sadeniemi 1951, Kiparsky 1968; see §4.1). The first two constraints, *heavy/W and *light/S, are categorical (as above). The final two, *heavier/W and *lighter/S, are their scalar counterparts, whose violations are real numbers generated by a phonetic function representing the weight percept. For the present purposes, this function is based solely on the log mean durations of rime types, as estimated from the auditory corpus above. In particular, for each locus of violation, the scalar constraint is violated by the locus's distance from the unmarked end of the scale (cf. McCarthy's 2003:82 paraphrase of HNUC). Here, the extrema are taken to be the minimum and maximum log mean durations ($\log(88\text{ms}) = 4.48$ and $\log(408\text{ms}) = 6.01$).[78] Consider, for example, the stressed rime *át* in candidate line (a). Because *át* occupies a weak position, it violates both *heavy/W (by 1) and *heavier/W (by $\log(264\text{ms}) - \log(88\text{ms}) = 1.099$). Note that the violation is the same if seconds or some other unit is used instead of milliseconds: $\log(.264\text{s}) - \log(.088\text{s}) = 1.099$. If a rime type is absent from the present data, the mean of its skeletal type is substituted (e.g. if *ép* is absent, the mean log duration of V̆C is used in its stead).

[78]Calibrating durations against the extrema has the potential virtue that the violation for a single locus is in practice bounded by the range $[0, c]$, where c is a smallish real, 1.53 here (though in principle, the maximum mean duration of a syllable type is unbounded). Other calibrations are conceivable. For example, one might use the grand mean instead of the extrema, subtracting or adding as necessary. On this approach, violations would drift into negative ('reward') territory for loci on the unmarked side of the mean.

	*heavy/W 1.348	*light/S 1.055	*heavier/W .449	*lighter/S .252	penalty	p
(a) paksumpi patsasta portin	1 (at)	0	1.10 (at)	.54 (or)	1.98	.06
(b) $s\widehat{ou}$ti seppo Ilmarinen	0	0	0	.80 (ep, il)	.20	.35
(c) asti nämät ovat hamin	0	3 (ä, o, a)	0	3.93 (ä, o, a)	4.16	0
(d) olla minä teːn hänelle	0	1 (i)	.20 (ä)	1.42 (i, eːn)	1.50	.08

Figure 107: Illustrative maxent tableau for the Kalevala meter.

Line (a) also contains one stressed syllable in a (non-initial) strong position, namely, *pór*. Although this is a heavy syllable, satisfying *light/S, it still receives a small penalty from *lighter/S for its distance from the heavy extremum (log(408ms)$-$ log(238ms) $=$ .539). Constraint weights (computed using Wilson and George 2008 with $\mu = 0$ and $\sigma^2 = 1000$) are given under their labels. To compute the penalty for line (a), its violation vector <1, 0, 1.1, .54> is multiplied by the constraint weight vector <1.348, 1.055, .449, .252>, and the result <1.348, 0, .49, .14> is summed, giving 1.98. The given p-value represents in this case the proportion of lines in the corpus with penalties higher than the candidate's. Thus, $p = 0$ for line (c) (from prose) indicates that no line of the Kalevala has that high a penalty. Line (d), while also from prose, is as harmonic as almost 10% of Kalevala lines by this metric. This cline of harmony recalls the notions of complexity (Halle and Keyser 1971) and gradient metricality (Hayes 1989b, Youmans 1989). Finally, note that this grammar is not a complete model of Kalevala weight mapping. For one, it ignores the fact that violations become increasingly costly towards the end of the line (§4.1.3); it also sets aside the apparently weaker constraints on unstressed syllables (§4.3.2).[79]

Though constraints such as (categorical) *heavy/W and (scalar) *heavier/W par-

[79]Increasing stringency can be modeled either by cloning the relevant constraints for each position or by including position as a variable in the constraint's violation function (i.e. definition), such that violations in later positions are amplified. Likewise, for unstressed syllables, one could either add constraints indexed to them (or to all syllables), or else build stress level into the violation function.

tially overlap in their work loads, their coexistence generates a typology in which categoricity can vary across systems. An independent argument for the simultaneity of categorical and gradient weight concerns meters such as the Homeric hexameter. Recall from §3.3 that the meter contains two position types, the longum and the biceps, both of which, if filled by a single syllable, permit only a heavy.[80] In this respect, the meter is categorical in its treatment of ligts and heavies, as could be modeled under the present theory by giving constraints such as *light/biceps and *light/longum sufficiently high weights.[81] At the same time, however, a clear weight gradient is observed within the heavies, such that the heavier the heavy, the more skewed it is towards the biceps, all else equal (§3.5). To model this difference, the weight of *lighter/biceps can exceed that of *lighter/longum, such that lighter heavies incur greater penalties in bicipitia.

Figure 108 is a maxent tableau for the hexameter. Since no phonetic corpus is available for Ancient Greek, for the sake of exposition, the Finnish data used above are also recruited here (as above, if a rime type is absent from the data, the skeletal mean stands in). Even though the categorical constraints are rarely violated, they need not be given the highest weights, since they overlap ('gang up') with their scalar counterparts in eschewing lights. Note that the grammar as stated is incomplete; for instance, it ignores the possibility of filling a biceps with a pair of lights, focusing only on the treatment of syllable weight.

---

[80]Exceptions can be found, especially in the first metron (fn. 28); elsewhere, they are rare. Exceptionality is generally trivial to model in maxent grammar, in which near-categorical constraints can be precisely weighted such that a suitably tiny proportion of the probability mass is allotted to such cases.

[81]To avoid confusion, I avoid labeling the hexameter position types strong and weak, though this is not to imply that they cannot be reformulated in such terms.

	*lighter /biceps 4.071	*light /biceps 3.557	*lighter /longum 3.522	*light /longum 3.324	penalty	p
(a) triplɛːj tetraplɛːj t' apo-tejsomen aj ke pot^{h}i zdews	.48	0	2.48	0	10.69	.53
(b) ej tar ^{h}o g' ewk^{h}ɔːlɛːs epi-memp^{h}etaj ɛːd' ^{h}ekatombɛːs	.38	0	2.53	0	10.46	.55

Figure 108: Illustrative maxent tableau for the Homeric hexameter.

In sum, categorization and a gradient phonetic interface are argued to interact in weight mapping. First, categorization can be incomplete, in which case polarization towards categories is evident but not so absolute that the grammar is insensitive to the underlying phonetic scale within categories. Second, even if categorization is complete, it can still be accompanied by intra-categorial gradience, as in the hexameter, which exhibits both a strict binary cutoff and sensitivity to a continuum of weight within the heavies.

10.2 On other possible approaches

As an alternative approach to gradient weight, consider first a theory under which weight is exclusively binary, but syllables are assigned to categories probabilistically, perhaps owing to variable mora projection (on non-variable mora-based weight, see Hyman 1985, Zec 1988, Hayes 1989a, Steriade 1991, Gordon 2002). For example, the coda of a V̆C rime might exhibit, say, a 90% chance of projecting a mora, while the second timing slot of VV might exhibit greater odds, such that the latter patterns as slightly heavier. More technically, the question is whether the grammar treats gradient weight as a continuous variable (as proposed in §10.1) or as a discrete (also known as Bernoulli) random variable.

The Greek hexameter furnishes an empirical argument against intra-heavy weight as variable binary weight. As discussed in §10.1, an intra-heavy scale is evident even among positions in which light syllables are categorically forbidden. In the fifth metron, for instance, the V̆C share of heavies is 69% higher in the longum than in the biceps, even though neither position can be filled by a light (West 1970). For an analysis in terms of variable binary weight, we can assume that progressively lighter heavies are progressively more likely to be categorized as light, while the lighter status of the longum is motivated by its being more tolerant of lights than the biceps. From this analysis, it follows that lighter heavies (e.g. V̆T and V̆N) should be found more than occasionally in light positions in the fifth metron, as should true lights (rime V̆) in the fifth longum. Both predictions are incorrect.

Such an approach would likely also treat timing slots individually, which I term ATOMIZATION, rather than considering properties of the whole rime as an unanalysed unit, as I did in §10.1. Even without reference to moras, one can imagine an atomized version of the approach in §10.1. In Tamil, for example, *glide-coda/W could outweigh *lateral-coda/W, motivating the lighter treatment of V̆L vis-à-vis V̆J. (For strong positions, in which V̆J is the less marked, *lateral-coda/S could outweigh *glide-coda/S.) The approach in §10.1 is simpler and more restrictive. First, under atomization, cases like Tamil would require (at least) several constraints to replicate the work done by a single constraint in §10.1, though it is difficult to say precisely how many. Is there a different constraint for each coda consonant? If not, how are consonants grouped? These sorts of questions are moot in §10.1. Indeed, the distributions in §9.1ff suggest that Tamil weight is continuous, rather than discretized into even several well-defined levels. One can always approximate a continuous distribution by superimposing increasingly fine-grained categories on it, though if the distribution reflects an intrinsically gradient variable, such an exercise is analytically extravagant.

Moreover, atomization risks massive overgeneration. For instance, one could generate Tamil′, a language phonetically identical to Tamil, except whose metrics treats V̆L as heavier than V̆J, reversing the weights of the aforementioned constraints. To avoid this problem, one might (1) stipulate a fixed ordering of constraints, (2) project constraint weights from a phonetic model, or (3) reformulate the constraints in terms of a stringency hierarchy. I address these three analytical tacks in turn.

First, fixing the ordering amounts to adding a set of stipulations that were unnecessary in §10.1. Furthermore, even with a fixed ordering, meters are generated in which position types exhibit unrelated patterns of conflation, e.g., one in which weight is strictly binary, but the cutoff falls between V̆L and V̆J for weak positions and between V̆T and V̆N for strong positions. Such a meter is compatible with the hierarchy T < L < J (etc.). In §10.1, by contrast, the binary criterion (if any) is constant (assuming that the shorthand predicates 'light' and 'heavy' are by definition complementary) and all scalar constraints access the same phonetic scale. See also de Lacy (2004) *contra* fixed rankings.

Second, one might project constraint weights from a phonetic model (cf. Steriade 2001b on projection in Optimality Theory), avoiding some of the aforementioned overgeneration problems (e.g. Tamil′) as well as the need for scalar constraints. However, the questions concerning the number and division of categories persist, as do the methodological concerns regarding the explosion of the constraint set in an attempt to simulate a continuous distribution (cf. Zhang 2007). Moreover, even insofar as this approach approximates a notational variant of §10.1, it is more complex on multiple dimensions, in that it requires both many more constraints and an additional grammatical component to project their weights.

Finally, expressing constraints in terms of a universal stringency hierarchy avoids some overgeneration issues (e.g. Tamil′) and arguably possesses theoretical and empirical advantages over fixed rankings (Prince 1999, de Lacy 2002, 2004). Since heaviness is marked in weak positions, for instance, a stringent weak-indexed constraint would penalize not only a particular rime, but any and all heavier rimes (e.g. *V̆L-or-heavier/W, *V̆N-or-heavier/W). Such constraints can be freely ranked, producing various patterns of conflation but never markedness reversals (as with Tamil′). Several of the problems outlined above also apply to stringency. For one, stringency overgenerates discrepant patterns of conflation for strong vs. weak positions. Further, it requires many more constraints than §10.1, including at least 18 for Tamil (nine levels in §7 times two position types), though perhaps many more, considering that the hierarchy in principle subsumes any weight distinction significant in any language.

The universal hierarchy is itself problematic. Consider, for instance, the fact that V̆R is lighter than V̆T in Kambaṉ (§2) but heavier than it in Homer (§3.2). If this difference is phonetically grounded, the approach in §10.1 gets it for free. A stringency approach, by contrast, can only accommodate such superficial reversals by adding complexity to the universal hierarchy (e.g. distinguishing between shorter and longer rhotics), further recapitulating the phonetics. Finally, the theory in §10.1 predicts that weight should reflect the phonetics not only in terms of the rank order of rime types, but also in terms of the magnitudes of various contrasts (e.g. in figure 74, V̆R < V̆T and V̆T < V̆N are both significant, though the magnitude of former difference is approximately twice as great as that of the latter). Because stringency specifies a rank order but is mute on relative differences, it is less restrictive. This final point also applies to any phonetic differences between languages that might be more fine-grained than a given theory of phonological features encodes (e.g. coda singleton [n] being slightly longer in one language than in another). Assuming such differences exist, the proposal here, unlike stringent constraints, predicts them to be relevant to gradient weight.

10.3 Local summary

This dissertation began by challenging the often assumed but rarely tested position that syllable weight is exclusively binary in metrics (see especially Jakobson 1933/1966 and Devine and Stephens 1976 for explicit statements of this view). Case studies of six quantitative meters drawn from three language families in every case reveal sensitivity to a continuum of weight within the heavies. Gradient weight in each case characterizes an interval scale, in which differences are a matter not just of ranking, but also of relative degree. Furthermore, the scales obtained across traditions are well correlated, not only with each other, but also with the hierarchies inferred from the crosslinguistic typologies of stress and other weight-sensitive phenomena. In §7 and §10.1, it is argued that the features of the gradient scale need not be stipulated, but emerge directly from the perceptual-phonetic interface via constraints whose degrees of violation are phonetically defined (as in Flemming 2001). The interaction of these scalar constraints with categorical constraints generates the attested typology of incomplete categorization and mixed categorical-gradient systems.

Part III
Gradient contributions of onsets to syllable weight

The preceding sections treated syllable weight solely as a function of rime structure. I now test whether properties of the onset also affect weight (of the syllable containing the onset). I first consider only the languages from the case studies above in which complex onsets are permitted, namely, Homeric Greek, Classical Latin, Old Norse, and Epic Sanskrit. Although the onset is irrelevant for categorical (heavy vs. light) weight in all these languages, I demonstrate that it contributes statistically to weight in all four, and always in the same way: The more segments in the onset (including one vs. zero), the greater the inferred aggregate weight of the syllable, judging by the same weight diagnostics previously employed.

I then turn to Tamil and Finnish, which forbid complex onsets. Considering only simple (one-consonant) onsets, I test for a correlation between onset duration and estimated metrical weight. In Tamil, a highly significant correlation is observed. In Finnish, the correlation is not significant in either the positive or negative direction. Once again, the conclusion is that characteristics of the onset matter for syllable weight, such that, just as with the rime, greater length and/or complexity correlates with greater weight.

11 Onset complexity and syllable weight

11.1 The Ancient Greek and Latin hexameters

As discussed in §3, a weight discrepancy between the hexameter longum (first half of the metron) and biceps (second half of the metron) permits the derivation of an intra-heavy scale of weight. In particular, because the biceps is underlyingly a heavier position type than the longum, heavies filling bicipitia tend to be heavier than heavies filling longa, such that the more skewed a heavy subtype is towards the biceps, the heavier it can be inferred to be. This method was used to extract a skeletal rime hierarchy of $\breve{\mathrm{V}}\mathrm{T} < \breve{\mathrm{V}}\mathrm{N} < \mathrm{VV} < \mathrm{VVC}$ (though the placement of VVC was problematic in the Latin; §3.6).

The same method can be exploited to investigate whether onset structure also affects weight. In this section, I divide onsets into three groups: null (i.e. an empty onset), simple (a single consonant), and complex (two or more consonants), abbreviated Ø, C, and C_2 respectively. Onset factors are entered into the model as a series of forward-difference coded fixed effects alongside the rimal factors. A regression table for Homeric Greek is given in figure 109. It is set up as in §3.5, with word shape as a random effect ($N = 68$; not shown) and forward-difference coded fixed effects (see §2.4.1). The outcome (dependent variable) concerns whether a syllable with the given features is placed in a biceps (coded 1) or longum (coded 0), correcting for possible confounds from position in the word and word template with the random effects, as before. One difference with respect to §3.5 is that only syllables immediately following heavies are taken as data in figure 109. This is to preclude possible interference from longer onsets being avoided after light syllables (e.g. if there are any aspects of (re)syllabification that are not covered by the conventional syllabification algorithm in §3.2, such that lights are gradiently heavier preceding longer onsets); recall that light syllables are found only in the biceps in the hexameter, not the longum,

		coefficient	standard error	t-value	p-value
intercept	(=$\breve{\text{V}}$T rime, Ø onset)	.4403	.0580	7.6	$< .00001$
rime $\breve{\text{V}}$N	[vs. $\breve{\text{V}}$T]	.0262	.0043	6.0	$< .00001$
rime VV	[vs. $\breve{\text{V}}$N]	.0357	.0035	10.2	$< .00001$
rime VVC	[vs. VV]	.0110	.0043	2.5	$= .01$
onset C	[vs. Ø]	.0372	.0067	5.5	$< .00001$
onset C_2	[vs. C]	.0123	.0051	2.4	$= .01$

Figure 109: The Greek model with three levels of onset complexity.

The significant, positive coefficients for both onset factors in figure 109 indicate that increasingly complex onsets are increasingly overrepresented in bicipitia (the heavier position type): $\text{Ø} < \text{C} < \text{C}_2$. The former contrast is highly significant while the latter is borderline.[82]

The same conclusion is reached if one leaves the rime-based factors out of the model, as in figure 110. This suggests that onset and rime structure are largely (though perhaps not entirely) independent of each other, as one might expect on typological grounds (Selkirk 1982, Kessler and Treiman 1997, Kapatsinski 2009).

		coefficient	standard error	t-value	p-value
intercept	(=Ø onset)	.4447	.0584	7.6	$< .00001$
onset C	[vs. Ø]	.0444	.0067	6.6	$< .00001$
onset C_2	[vs. C]	.0128	.0050	2.5	$= .01$

Figure 110: The Greek model with onset complexity only.

Note that the avoidance of null onsets in the biceps relative to filled ones cannot be attributed to a confound from (gradient or categorical) shortening of the preceding

[82]The latter contrast is borderline at $p = .01$ if one takes into account the Bonferroni correction (§3.5.1). However, one could perhaps make the case that a one-tailed p-value is appropriate here, if the hypothesis is taken to be unidirectional — ‘a complex onset contributes more to weight than a simple one’ — in which case $p = .007$, under the corrected criterion.

vowel in hiatus. Given that vowels tend to (if anything) shorten in hiatus (known as CORREPTION), not lengthen (though cf. Kiparsky 1968:142), null onsets might be expected to be better represented in the biceps, as shorter vowels are more acceptable in the longum than in the biceps (§3.3). The fact that the opposite situation is found — null onsets being better represented in the longum — would then seem to be motivated by the intrinsic contribution of a filled onset to weight rather than to the effect of a null onset on vowel length.

An additional piece of evidence likewise suggests that the presence of an onset per se, and not an effect on vowel length, is at least in part driving the $\emptyset < C$ onset contrast in Ancient Greek. As Maas points out, Callimachus and Nonnus in their hexameters permit hiatus at the longum-biceps juncture only if the biceps is filled by a pair of lights (Maas 1962:89). Maas does not mention Homer in this regard, but Homer appears to exhibit the same tendency. As figure 111 illustrates, in biceps-initial position, null-onset heavies are much more underrepresented than null-onset lights (though both are underrepresented). Thus, not only are null onsets avoided in the biceps, but they are avoided more stringently when the biceps is filled by a single syllable (Fisher's $p < .0001$). A potential explanation of this discrepancy is that a single heavy in the biceps (the heavier half of the metron) is not heavy enough if it is onsetless, whereas when the biceps is filled by a pair of lights, the constraint is moot, or less applicable (cf. West 1982 on a pair of lights patterning as heavier than a single heavy). At any rate, this discrepancy is unlikely to be driven by effects of hiatus on the preceding (longum) vowel, since longum C_0VV is equally in hiatus before both biceps VV and biceps $\breve{V}.C_1\breve{V}$.

	biceps-initial	elsewhere	% biceps-initial
Ø-onset light	1,126	3,538	24.1%
Ø-onset heavy	210	7,483	2.7%

Figure 111: A closer look at the distribution of empty onsets in Homer.

The contrast between null and filled onsets, with the latter patterning as heavier (i.e. more biceps-attracted), is also found in Vergil's hexameter, though the contrast between C and C_2 found in Homer is nonsignificant in Vergil. A regression table (showing only the relevant onset factors) is given in figure 112. As with the Greek above, only post-heavy heavies are included as data.

		coefficient	standard error	t-value	p-value
intercept	(=Ø onset)	.3566	.0582	6.1	$<$ .00001
onset C	[vs. Ø]	.1490	.0183	8.1	$<$ .00001
onset C_2	[vs. C]	$-$.0032	.0070	$-$.5	= .6

Figure 112: Onset weight in the Latin hexameter.

The findings for onset weight in the Greek and Latin hexameters are summarized as Hasse diagrams in figure 113.

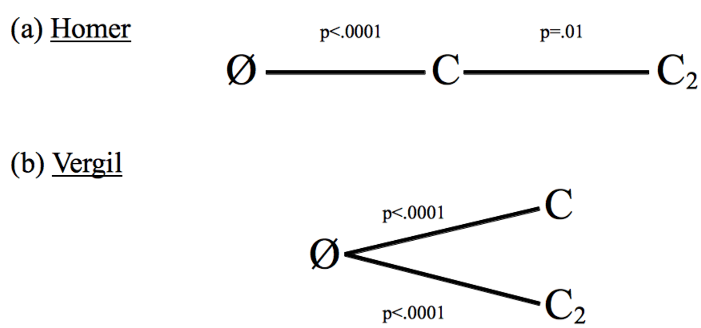

Figure 113: Hasse diagrams for onset weight in Homer (top) and Vergil (bottom).

11.2 The Old Norse dróttkvætt

Greater complexity also correlates with greater weight in the Old Norse *dróttkvætt*. In §6, two independent tests revealed a several-tiered scale of weight (specifically, considering only the skeletal structure of the rime, $\breve{\mathrm{V}} < \breve{\mathrm{V}}\mathrm{C} < \mathrm{VV} < \{\breve{\mathrm{V}}\mathrm{CC}, \mathrm{VVC}\} < \mathrm{VVCC}$, one based on stressed syllables (which, as they get heavier, skew more strongly towards strong as opposed to weak positions) and one based on unstressed syllables (which, as they get heavier, are progressively more avoided in cadential weak positions relative to non-cadential ones). I now employ both of these same tests to demonstrate that increasing onset complexity also significantly contributes to weight. Since onsets are on average more complex in Old Norse than in Latin or Ancient Greek, I consider four levels of onset complexity in this section, namely, Ø (null), C (simple), CC (two consonants), and CCC_1 (three or more consonants). Although two syllabification algorithms — onset and coda maximization — were considered in §6, in this section, the tests assume only the former; under the latter, as defined in §6, complex onsets do not exist.

The regression table for the first diagnostic, stressed syllables, is given in figure 114. Only onset factors are shown in the table. The hierarchy is monotonic: $\text{Ø} < \mathrm{C} < \mathrm{CC} < \mathrm{CCC}_1$ (all $p < .005$). In other words, the longer the onset of a stressed syllable, the more likely that syllable is to be placed in a strong position. Note that because strong positions usually follow weak ones in the *dróttkvætt* (§6), this preference for longer onsets in strong positions cannot be attributed to a confound from the weight preference of the preceding position, (which would, if anything, pressure for a shorter following interlude, being preferentially light).

		coefficient	standard error	z-value	p-value
intercept	(=Ø onset)	$-.8842$	.6602	-1.3	$= .2$
onset C	[vs. Ø]	.2683	.0433	6.2	$< .00001$
onset CC	[vs. C]	.2012	.0442	4.5	$< .00001$
onset CCC_1	[vs. CC]	.5960	.1942	3.1	$= .002$

Figure 114: Onset weight in stressed syllables in the *dróttkvætt*.

Figure 115 is the regression table for the second diagnostic, unstressed syllables. Despite the completely disjoint set of data from the previous test, the same result is obtained: Ø $<$ C $<$ CC $<$ CCC_1 (all $p < .005$). In this case, the longer an unstressed syllable's onset, the more that syllable is eschewed from cadential weak positions in favor of less stringently regulated weak positions earlier in the line (§6).

		coefficient	standard error	z-value	p-value
intercept	(=Ø onset)	.968	1.157	.8	$= .4$
onset C	[vs. Ø]	1.911	.317	6.0	$< .00001$
onset CC	[vs. C]	1.012	.050	20.2	$< .00001$
onset CCC_1	[vs. CC]	.748	.247	3.0	$= .003$

Figure 115: Onset weight in unstressed syllables in the *dróttkvætt*.

The scale from both tests is summarized as a Hasse diagram in figure 116.

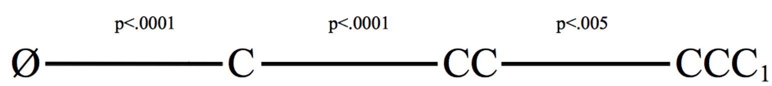

Figure 116: Hasse diagram for onset weight in Old Norse.

11.3 The Epic Sanskrit šloka

11.3.1 Onset weight in Sanskrit: line-initial syllables

I begin with the set of line-initial syllables in Epic Sanskrit (§5). This is an expedient place to start for two reasons. First, there is no need to worry about controlling for the heaviness propensity of the preceding position, as there is no (relevant) preceding metrical position. Second, there is no question of how the intervocalic interlude is divided between coda and onset, since there is no intervocalic interlude. Sanskrit onsets range from zero to four consonants. Four is exceedingly rare (five tokens in the corpus, all *stry*), while three is somewhat rare (0.1% of tokens). I therefore distinguish three levels of onset complexity in this section: null (Ø), simple (C), and complex (C_2, i.e. two or more consonants).

As explained in §5, the line-initial position of the *śloka* is always anceps, provided the proper conception of ANCEPS as being a position in which both heavies and lights are welcome, but which might nevertheless exhibit a preference, however weak, for one or the other. Indeed, the heaviness of the line-initial anceps is often significantly different from chance. For example, consider lines beginning with a heavy-final bisyllabic word, #XH#. In the corpus as a whole, X in XH words is heavy 71% of the time. But as the opening, X in XH is heavy only 58% of the time ($p < .001$). We thus observe a small but significant preference, above and beyond lexical statistics, for an iambic as opposed to spondaic opening when the first foot is a heavy-final word. In the same fashion, I compute the expected heaviness proportion of the first position of every line-template in the corpus. The observed proportion is divided by the expected proportion to estimate the position's underlying heaviness propensity. Figure 117 offers a few illustrations.

	template of initial $\sigma\sigma\sigma\sigma$	X=heavy observed	X=heavy expected	O/E	poets' preference
(a)	XLH#H...	70.0%	49.7%	1.41	prefer heavies
(b)	XHLL#...	53.0%	55.1%	.96	neutral
(c)	XHL#H...	50.4%	75.1%	.67	prefer lights

Figure 117: The prosody of the opening: observed vs. expected examples.

If onset complexity affects weight, we would expect heaviness propensity to be positively correlated with onset complexity, all else being equal. This is indeed the case, as figure 118 reveals. As before, the model is a mixed model, with word position/shape as a random effect. Rime shape (in this case, $\breve{V}$, $\breve{V}C$, $\breve{V}C_2$, VV, and VVC_1) is also included in the model, though omitted from the figure. In this case, the model is not forward-difference coded, but rather dummy coded (§2.4.1). This means that both factors are interpreted with respect to the intercept. C (simple) is used as the intercept, both because it is the intermediate level and because it is by far the most frequent level (94% of tokens). The outcome is the estimated heaviness propensity of the initial position in the context of its line.

	coefficient	standard error	t-value	p-value
onset Ø	$-.0319$	.0005	-66.7	$< .00001$
intercept (=C onset)	1.0208	.0094	109.0	$< .00001$
onset C_2	.0043	.0005	8.8	$< .00001$

Figure 118: Onset complexity in line-initial Sanskrit syllables.

We therefore observe the onset hierarchy Ø $<$ C $<$ C_2 (all contrasts $p < .0001$). Although figure 118 subsumes both heavies and lights, we might also consider the two categories separately to see whether parallel effects obtain. I therefore split figure 118 into two models, treating lights in figure 119 and heavies in figure 120.

	coefficient	standard error	t-value	p-value
onset ∅	$-.0303$	.0008	-38.5	$< .00001$
intercept (=C onset)	1.0139	.0080	126.8	$< .00001$
onset C_2	.0101	.0008	12.6	$< .00001$

Figure 119: Onset complexity in line-initial Sanskrit syllables (lights only).

	coefficient	standard error	t-value	p-value
onset ∅	$-.0338$	.0006	-56.3	$< .00001$
intercept (=C onset)	1.0231	.0104	98.0	$< .00001$
onset C_2	.0001	.0006	0.3	$= .8$

Figure 120: Onset complexity in line-initial Sanskrit syllables (heavies only).

For lights, the hierarchy $\emptyset < C < C_2$ (all $p < .0001$) is once again observed, as it was in the aggregate. For heavies, $\emptyset < C$ once again obtains, while $C < C_2$ is nonsignificant. In sum, first, the absence vs. presence of an onset is a stronger contrast in both lights and heavies than the contrast between simple and complex onsets (see also Gordon 2005); this was also observed in Latin in §11.1, in which $\emptyset < C$ was significant while $C < C_2$ was not. Second, $C < C_2$ is significant only among lights in the Sanskrit.

I offer two (not incompatible) possible explanations for this discrepancy between heavies and lights. First, it might be due to a kind of Weber's law effect in weight perception. For light syllables, a small change in weight (such as that induced by onset complexity) might be more perceptible than a change of the same absolute magnitude to an already heavy syllable (cf. Lunden's 2006 proportional increase theory of weight). As a second possibility, research on the perceptual centers (P-centers, or perceived/produced downbeats) of syllables has shown that P-centers tend to lag as rime length increases (Marcus 1981). In other words, in a heavier syllable, the P-center will tend to be further displaced from the beginning of the onset than it would be in a light syllable with the same onset. If onset material following the P-center is perceived to be more relevant to syllable weight than onset material preceding the P-center, this would also motivate a discrepancy in onset weight sensitivity between light and heavy syllables.

11.3.2 Onset weight in Sanskrit: line-medial syllables

I now turn to line-medial syllables in Sanskrit. This situation is more complex than that of §11.3.1 because we are now dealing with a consonantal interlude straddled by two metrical positions, both of which might simultaneously influence its shape. For example, consider the interlude between a heavy position and a light position. Even if the light position pressures for a shorter onset (and hence shorter preceding interlude), the heavy position might counteract this pressure by exerting its own preference for a longer following interlude. I am not claiming that this is in fact a confound, or that it is somehow predicted by standard approaches to syllable division, only that it is a possible confound that should be controlled.

A simple way to control for influence from the preceding position is to hold it constant. In this case, I include a syllable as data only if it immediately follows a crucially heavy position, which I take to be a ≥99.9% heavy position (in the rare case that a light occupies such a position, the datum is excluded). Just over a million such syllables are found in the corpus. Fortunately, the metrical positions immediately following crucially heavy positions cover the range of weight preferences fairly evenly, from crucially light (31% of post-crucially heavy positions), through various shades of anceps (44%), to crucially heavy (25%). I run the same model as in §11.3.1, with the same controls, on this new data set. As in §11.3.1, the dependent variable is taken to be the heaviness propensity of each position. Figures 121 and

122 are for lights and heavies, respectively.

	coefficient	standard error	t-value	p-value
onset Ø	$-.1993$	.0144	-13.8	$< .00001$
intercept (=C onset)	.5648	.1432	3.9	$< .0001$
onset C_2	.0290	.0019	15.3	$< .00001$

Figure 121: Onset complexity in line-medial Sanskrit syllables (lights only).

	coefficient	standard error	t-value	p-value
onset Ø	.0096	.0071	1.4	$= .18$
intercept (=C onset)	1.1266	.0406	27.7	$< .00001$
onset C_2	.0009	.0005	2.0	$= .05$

Figure 122: Onset complexity in line-medial Sanskrit syllables (heavies only).

Once again, Ø < C < C_2 (all $p < .0001$) emerges among lights. Among heavies, however, no contrast achieves significance (though C < C_2 comes close at $p = .05$). Figure 123 summarizes the findings for Sanskrit. A check indicates significance, while a blank indicates no significant contrast in either direction. In none of the four contexts does a significant contrast contradict a significant contrast found elsewhere.

	Ø < C	C < C_2
line-initial lights	√	√
line-medial lights	√	√
line-initial heavies	√	
line-medial heavies		

Figure 123: Summary of onset effects for Sanskrit.

A final clarification is in order about the line-medial results. I have thus far employed an algorithm for parsing onsets out of complex interludes in Sanskrit which might be called RESTRICTED ONSET MAXIMIZATION. It is onset maximization in the sense that priority is given to making onsets as complex as the phonotactics (evidenced largely by word-initial sequences) permits (§5; cf. Hermann 1923:257ff, Devine and Stephens 1994:41ff, Kessler 1998). It is restricted because there is one context in which maximization is forbidden, namely, when the onset would consume an entire complex interlude. For example, even though *kr* is a licit onset, *cakram* 'chakra' can only be parsed as *cak.ram*, never **ca.kram*, as is evident from the fact that its initial cannot be placed in a breve. Thus, parses such as **ca.kram* and **a.stram* 'missile' are ruled out on metrical grounds. But when it comes to, say, *saṃskṛtam* 'Sanskrit', the metrics provides no obvious guidance on whether the boundary precedes or follows medial *s*; either way, the initial comes out heavy. These sorts of ambiguities in parsing complex interludes raise the issue that the algorithm assumed here for Sanskrit might be flawed, skewing the results in this section.

However, even if details of the algorithm are wrong, the main finding, that prenuclear material aggregately influences the distribution of the following syllable, is robust. Just as more material in the rime adds weight, more material in the onset has a parallel effect. A priori, three elements can be arranged six ways, but the same hierarchy $\emptyset < \mathrm{C} < \mathrm{C}_2$ emerges in two independent tests, every link highly significant. Even if the parsing algorithm is imperfect, sometimes overparsing or underparsing onsets (or being overly discrete or deterministic), the present findings are still conclusive in certain respects. For one, if we take as a null hypothesis the claim that onset complexity has no bearing on the metrical distribution of syllables, the present results suffice to reject this hypothesis, even if they leave some details of implementation unresolved. Moreover, the same results were obtained for line-initial syllables, for which there is no question of how interludes are divided between syllables. The same could be said for medial $\emptyset < \mathrm{C}$, assuming *VC.V. In other words, if one assumes *VC.V (cf. Peperkamp 1997:§2, Blevins 2003, and references therein), for four out of the five check marks in figure 123, the syllabification algorithm is moot.

12 Onset duration and weight

12.1 Tamil: Kambaṉ's epic

Tamil, as discussed in §2, lacks complex onsets. Nevertheless, one can still investigate whether syllables with (on average) phonetically longer onset consonants are treated as statistically heavier than syllables with typically shorter ones. In other words, instead of using segment count as a metric for onset length, as in the preceding sections, one can employ phonetic duration as a metric, even while restricting one's attention to simple onsets. I demonstrate here that onset duration is strongly correlated with syllable weight in Tamil metrics, such that syllables with durationally longer onsets are more attracted to strong positions (and more avoided in weak ones) than syllables with shorter onsets.

A phonetic corpus of 351 Tamil syllables was described in §9.1. In addition to measuring acoustic features of the rime, the duration of each onset (if any) was measured. In this section, I consider only simple onsets (i.e. onsets comprising exactly one consonant) immediately following a vowel. There is therefore no issue of dividing geminates here; all onsets used as data are intervocalic singletons. Figure 124 plots the duration (ms) of each post-vocalic onset consonant token (x-axis) against the mean estimated metrical weight of all syllables containing that onset (also in postvocalic position) in Kambaṉ's epic (y-axis). More precisely, metrical weight is derived from the linear propensity model in §2.5ff. Recall that this model provides an estimate as a log-odds ratio of the strength of every position in the corpus, taking into account the position's context. In the present case, for each onset consonant, the log-odds ratios of all syllables with that onset are averaged, giving a mean estimated heaviness propensity of the onset. The greater this propensity, the more skewed the onset is towards metrically stronger positions.[83] The horizontal stratification visible in the figure is due to multiple phonetic tokens being mapped to the same weight estimate. The correlation (depicted by the red regression line) is positive and significant (Pearson's $r = .71$, Spearman's $\rho = .73$, both $p < .00001$).

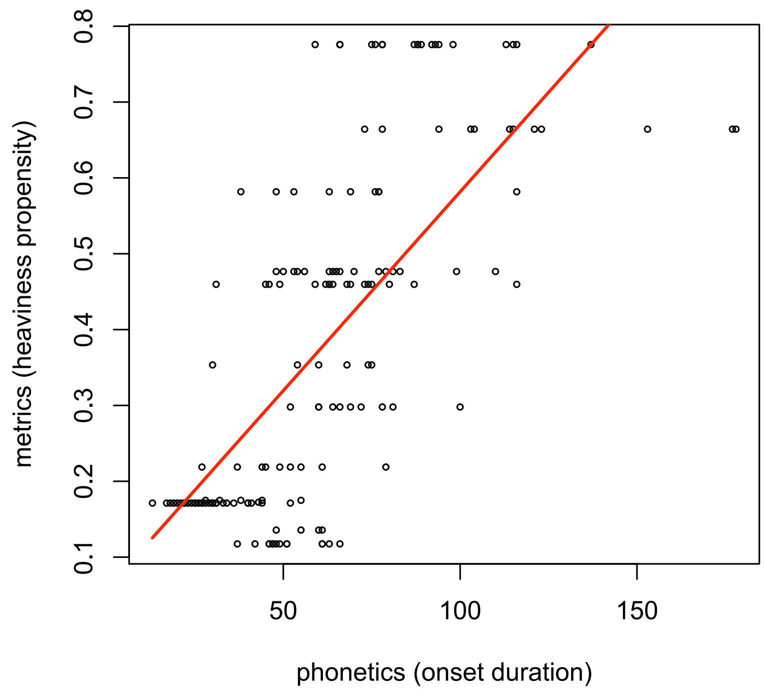

Figure 124: Onset duration (tokens) (x-axis) vs. metrical weight (y-axis).

Figure 125 is based on the same data as figure 124 but shows the mean duration of each onset type in the phonetic data instead of the individual durations of tokens separately. For example, 'R' (i.e. [ɾ]) is the shortest onset consonant in the phonetic

---

[83]Unlike some previous plots, the metrical propensities here are average log-odds, not proportions or probabilities. It is merely a coincidence that their range is so close to $(0, 1)$.

data (leftmost in the plot); it is also among the most weak-skewed onsets in the metrical corpus (near the bottom of the plot). The figure depicts only onset types with at least five (post-vocalic) tokens in the phonetic data. The size of each letter is proportional to its frequency in the phonetic data. With this greater degree of abstraction, the correlation is somewhat tighter, though the p-values are less stringent (though still well within significance) due to the fewer data points (Pearson's $r = .84$, $p = .001$; Spearman's $\rho = .84$, $p = .003$). IPA translations of the letters in the plot are as follows: r = [ɾ], R = [d], L = [ɭ], T = [ɖ], k = [g] or [ɣ], t = [d̪], l = [l], v = [ʋ], n = [n] or [n̪], m = [m], and c = [c] or [s]. Voicing is not contrastive, and the letters romanized as voiceless here (such as *t* and *T*) are usually voiced (as singletons) intervocalically, the only context under consideration here.

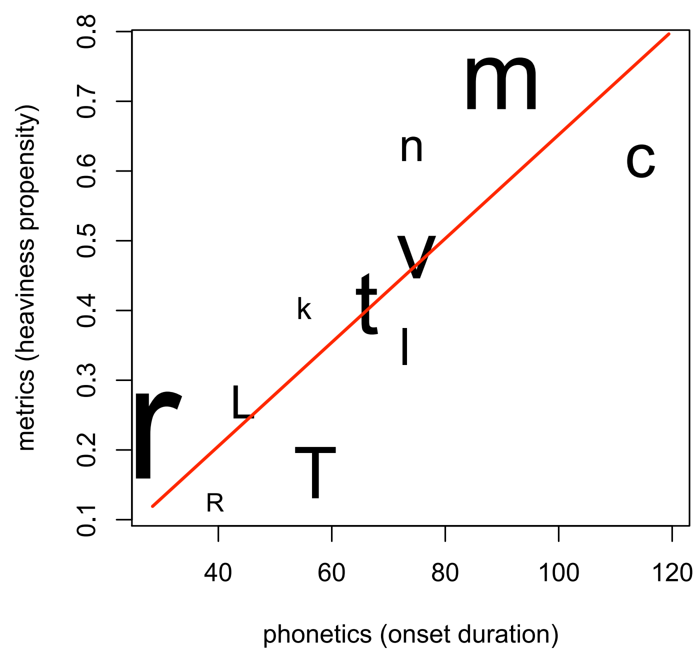

Figure 125: Onset duration (types) (x-axis) vs. metrical weight (y-axis).

12.2 Finnish: the Kalevala

The correlation of onset duration and syllable weight (as inferred from metrical diagnostics) is now tested for the Finnish Kalevala meter (§4) using the same method as just employed for Tamil in §12.1. The token (left) and type (right) correlations are plotted in figure 126 (cf. figures 124 and 125). The y-axis in both plots is the proportion of stressed syllables with each onset in strong as opposed to weak positions. The x-axis is the duration (ms) of the onset token (for tokens) or mean duration of onset tokens (for types), based on the phonetic data in §9.3. As in §12.1, only immediately postvocalic onsets are considered in both the phonetic and metrical data. In the plot for types, only types with frequencies of at least 5 are shown. None of the correlations is significant (all $p > .5$).

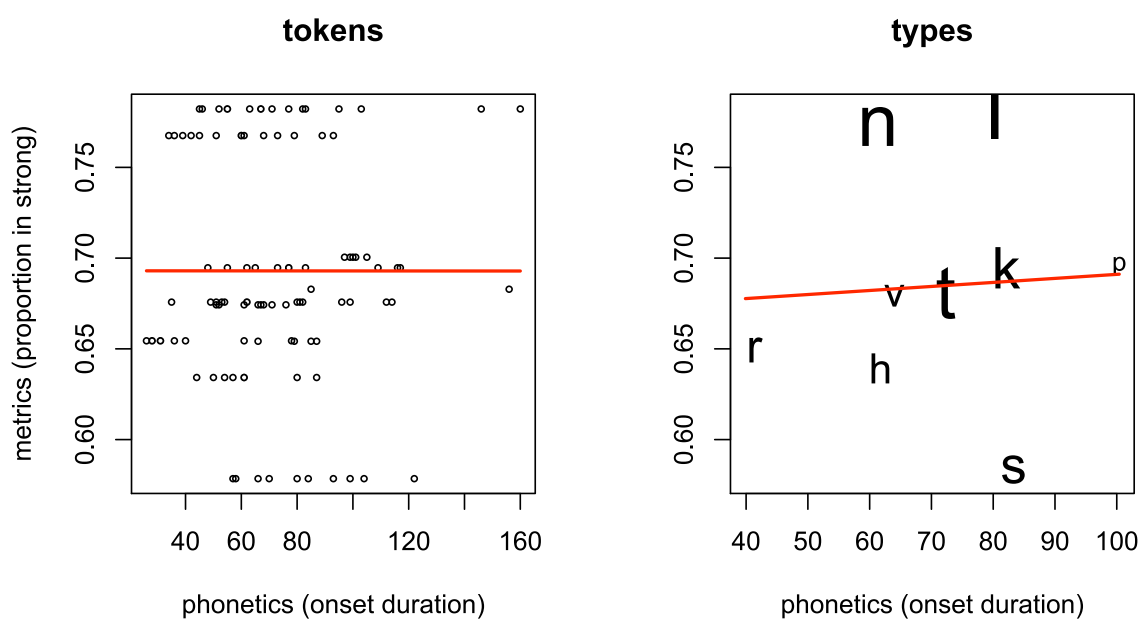

Figure 126: Onset duration (x-axis) vs. metrical weight (y-axis) in Finnish.

This non-result is perhaps not surprising, given that the Finnish meter is apparently much less sensitive than the Tamil to gradient weight effects in general (see figure 106). Furthermore, one possible problem with the present data is that only the onsets of peninitial syllables were measured in the Finnish data (originally in order to test whether interval duration provides a superior fit to rime duration in §9.2). In the present metrical data, however, only stressed syllables are considered in computing metrical strengths (see §4.1), such that only word-initial onsets are relevant.

13 Onset weight: summary

In §11.1–§12.2, I demonstrated that onset structure significantly affects the placement of syllables in a variety of quantitative meters. Most generally, all else being equal (e.g. controlling for word shape and position in the word), the longer the onset, whether in terms of segmental complexity or duration, the greater its skew towards stronger positions. This correlation between onset size and metrical weight supports the onset as an intrinsic factor in syllable weight. As far as I am aware, onset effects on weight have not been previously reported for any of the languages treated here.

Figure 127 summarizes the results for the six metrical corpora treated in the preceding sections. A check indicates a significant result for the test labeled at the top of the column (see above for exact values). If the contrast was tested but not significant, the cell indicates '(nonsig.)'. None of the contrasts was ever significant in the opposite direction in any of the tests (e.g. there was no case in which C $<$ Ø). Due to the varying natures of the phonologies and meters, not every contrast was tested in every language. In particular, two of the languages (Tamil and Finnish) lack complex onsets, the ancient languages lack phonetic data, and in some of the languages (Greek, Latin, and Sanskrit) CCC_1 onsets are not frequent enough to justify separating from CC. Contrasts that are not applicable are marked as ' untested '.

corpus	Ø < C	C < CC	CC < CCC	C duration ∼ weight
Greek (Homer)	√	√	untested	untested
Latin (Vergil)	√	(nonsig.)	untested	untested
Old Norse (stressed)	√	√	√	untested
Old Norse (unstressed)	√	√	√	untested
Sanskrit (initial lights)	√	√	untested	untested
Sanskrit (medial lights)	√	√	untested	untested
Sanskrit (initial heavies)	√	(nonsig.)	untested	untested
Sanskrit (medial heavies)	(nonsig.)	(nonsig.)	untested	untested
Tamil (Kamban̠)	untested	untested	untested	√
Finnish (Kalevala)	untested	untested	untested	(nonsig.)

Figure 127: Summary of onset effects.

From this table, it is evident that both Ø < C and C < CC are robust results across the languages, in that they are significant in most of the tests, and always in the same direction, never in the opposite direction. In Old Norse, in which highly complex onsets are particularly common, CC < CCC is also significant in two independent tests. Finally, the rightmost column ('C duration ∼ weight') refers to the test of correlation between the phonetic durations of simple onsets and metrical weight. In Tamil, this correlation was relatively high (always over 0.7 by various tests) and highly significant (always $p < .0001$). In Finnish, however, no significant correlation obtained, positive or negative.

Onset complexity has been documented as a factor in syllable weight in the stress systems of several languages, the most famous being Pirahã, in which, among other distinctions, VV < CVV, i.e. Ø < C (Everett and Everett 1984, Everett 1988). In fact, Gordon (2005) reports that both Ø < C and C < C_2 are found as weight criteria in stress (cf. Davis 1988, Goedemans 1998). By my count, Gordon (597ff) offers nine examples of languages in which the presence vs. absence of an onset is at least one of the weight distinctions, including Alyawarra, Arrernte, Banawá, Iowa-Oto, Júma, Lamalama (actually a small group of languages), Manam, Mbabaram, and Pirahã (see Gordon for references). He cites two languages observing C $<$ C_2 as a criterion, namely, Bislama and Nankina (though cf. Topintzi 2010 for possible reanalyses). Gordon does not cite Kelly (2004), whose article is roughly contemporary with his own, but English can now also be added as a language observing (statistical) weight contrasts between both onset Ø $<$ C and C $<$ C_2 in its stress system. The following sections (§14ff) further support the relevance of onset complexity to syllable weight in English.

Topintzi (2010) (see also Topintzi 2005, Topintzi 2008) argues for intrinsic onset weight from not only stress but several phonological phenomena, including compensatory lengthening (cf. Kiparsky forthcoming), word minimality, and geminate onsets. Yet another domain in which onset complexity arguably affects weight is that of binomial ordering, e.g. *fair and square* or *sea and ski*, where the second element might be more likely to have a complex onset than the first, all else being equal (e.g. Cooper and Ross 1975, Benor and Levy 2006). Finally, it has been established in the P-center literature that, in general, adding more material to the onset causes the downbeat to be perceived and produced earlier in the syllable (Morton et al. 1976, Marcus 1981, *et seq.*).

In conclusion, the crosslinguistic typology provides convergent support for onset structure as a factor in weight from a variety of domains. I have added new types of evidence with the present studies of the metrical impacts of onset complexity and onset duration revealing that even in languages in which onset complexity plays no role in categorical weight, its gradient effects can be inferred from the poets' treatment of syllable types in corpora. Note, finally, that any progressive onset weight effect (such as all the ones documented here) is problematic for a vowel-to-vowel interval account of rhythmic constituency in meter (see §9.2 for references and discussion). Under such an approach, an onset is never parsed into the same interval as an immediately following nucleus. If one assumes that the weight of an interval is determined by its contents, progressive onset weight effects would be unexpected.

Part IV
Gradient weight in English stress

14 Introduction

Just as in the case studies of quantitative meters in part I, stress systems in which stress placement exhibits variation (at least in some contexts) permit one to test whether speakers are sensitive to factors in weight that are more fine-grained than a binary heavy/light distinction, with the expectation that progressively heavier syllable types will be (if anything) progressively stronger attractors of stress. For background, I first briefly discuss categorical quantity-sensitive stress systems, such as Yana, Kashmiri, and Pirahã. I then turn to the somewhat different situation in the stress system that is the focus of this chapter, namely, English.

Quantity-sensitivity in stress allows (by definition) the diagnosis of at least one weight distinction. For example, in Yana, the first heavy syllable in the word receives stress. In all-light words, the first syllable receives stress (more succinctly, 'stress leftmost heavy, else leftmost'; Sapir and Swadesh 1960, Gordon 2002). Figure 128 provides two examples, each with an ungrammatical comparison. This system, being entirely categorical, diagnoses a binary criterion for weight: A light syllable has the form $C_0\breve{V}$; other syllables are heavy. In such a system, there is no sense in which $C_0\breve{V}C$ is lighter than C_0VV, or in which the difference between light and heavy or any other grades is quantifiable.

(a)	sibúmk'ai	'sandstone'	(*síbumk'ai)
(b)	p'údiwi	'woman'	(*p'udíwi)

Figure 128: Quantity-sensitive stress in Yana.

Categorical systems can also diagnose more than two degrees of weight. Kashmiri, for instance, distinguishes three: $C_0\breve{V} < C_0\breve{V}C < C_0VV$ (e.g. Morén 1999, 2000, and references therein). The final syllable of a Kashmiri word is extrametrical (except in monosyllables, in which it is stressed). Among non-final syllables, the leftmost syllable of the heaviest category available receives the stress. Pirahã stress observes an even more articulated hierarchy of five levels: $G\breve{V} < K\breve{V} < VV < GVV < KVV$ (where G is a voiced consonant a K a voiceless one; Everett and Everett 1984, Everett 1988, Gordon 2005). In Pirahã, within a final three-syllable window, the rightmost syllable of the heaviest category available receives stress. (See also Central Alaskan Yup'ik, Woodbury 1987; Nanti, Crowhurst and Michael 2005; and various languages in de Lacy 2004.)

All the aforementioned systems are described as being categorical, in the sense that they are deterministic: For a word with a given structure, only one possible stress assignment is licit. On the other hand, in many languages, stress is described as being 'free' or 'mobile', in the sense of being unpredictable and therefore lexically listed. Still other languages fall somewhere between these two extremes, with stress being largely predictable, or deterministic in a number of contexts, but with variation or arbitrariness being possible in some contexts (the more marked options sometimes being regarded as listed exceptions to an otherwise regular system of defaults). English stress (e.g. Chomsky and Halle 1968, Hayes 1982, Halle and Vergnaud 1987, Kager 1989, Burzio 1994, McCarthy and Prince 1995, Pater 2000, Carpenter 2010, etc.) is such a system.

Consider the English examples in figure 129. Pair (a) illustrates fixed initial and fixed final stress in disyllabic nouns of the frame C[o͡ʊ]C[iː] (stress judgments here follow *The American Heritage Dictionary*, Pickett 2000). To highlight the arbitrariness in these cases is not to deny that one of the two patterns might be less marked than the other nor to imply that such differences lack etymological explanations.

They merely illustrate that from a synchronic perspective, both patterns are possible (cf. Pater 2000). Similarly, in (b) and (c), nouns of the frame C_1[o͡ʊ]C[e͡ɪn] can be initially, finally, or variably stressed (the details in suites like this one vary by dialect, which only reinforces their synchronic arbitrariness). Stress in disyllabic verbs is also largely but not entirely predictable in English. Consider the two verbs of the frame [ɪn]C[ɹ̩] in (c). Of course, English stress is also rife with generalizations of varying productivity. The point here is simply that these generalizations do not add up to a fully deterministic system. The existence of regions of unreliability permits one to gauge the statistical strengths of various predictors of stress assignment.

(a)	dhoti	/dó.ti/	goatee	/go.tí/
(b)	propane	/pɹó.pen/	cocaine	/ko.kén/
(c)	ptomaine	/tó.men/	∼	/to.mén/
(d)	injure	/ín.ʤɹ̩/	infer	/ɪn.fɹ̩́/

Figure 129: Some points of arbitrariness in English stress assignment.

In this section, I focus on the influence of syllable structure in stress placement in monomorphemic disyllables, first in the lexicon and then in wugs (i.e. experimental non-words, Berko 1958). I argue from both phenomena that English stress is sensitive to a gradient continuum of weight in addition to the traditionally acknowledged heavy/light distinction. Moreover, in both paradigms — stress in the lexicon and in wugs — both the structure of the rime and the structure of the onset (as in part III) are significant factors in weight.

15 Gradient weight in the lexicon

As a corpus of existing English words, I take 7,930 monomorphemic (or highly lexicalized) disyllables from the MRC Psycholinguistic Database (Coltheart 1981) and CELEX (Baayen et al. 1995). Words from these two sources with the requisite features were combined into a single list without duplicates. It is difficult in many cases to draw a precise line between morphologically simple and complex words without operationalizing complexity using experimentally derived norms, and even then any binary cutoff would be largely arbitrary. For the present purposes, I inherit complexity judgments from (1) the MRC database, whose entries come annotated for morphological status, and (2), for words not in the MRC database, from annotations of productive affixes and compounds in CELEX by Bruce Hayes.

The data are also coded for part of speech, in this case making only a binary distinction between NOMINAL (i.e. noun or adjective) and VERBAL (i.e. verb or adverb). Other categories, such as interjections, are excluded. For the purposes of predicting stress assignment in English, this binary classifier does not perform significantly worse than one with four levels (noun, adjective, verb, adverb) in terms of classification accuracy.[84] Each word is also annotated for whether its primary stress is initial or final. Secondary stress (if any) is ignored.

15.1 Rime structure

I begin by testing the influence of coda complexity on stress assignment in nominals (setting aside verbals for the moment). A simple logistic model is summarized in figure 130, showing that coda complexity exhibits a positive and significant correlation with primary stress assignment. In other words, the more complex the coda of a syllable, the more likely that syllable is to bear primary stress. Each syllable is a datum. Because the corpus contains 5,878 disyllabic nominals, the present model is

[84]Specifically, in the present corpus of disyllables, nouns and adjectives receive initial stress 84% and 85% of the time, respectively, while for verbs and adverbs, the respective rates are 32% and 43%, respectively. Given the frequencies of these categories, neither difference is statistically significant.

based on 11,756 (5,878 × 2) data points, one for each syllable. Each datum is coded for whether it is primary stressed (1) or not (0). This variable is the outcome being predicted by the model.[85] Thus, any significant positive coefficient in the model increases the likelihood that a syllable bears stress. CODA SIZE is the number of consonants in the coda (zero to three in the present data). INITIALITY refers to whether the syllable is initial or final in the disyllable, effectively controlling for a possible confound from position in the word. Nominals have a strong tendency for initial stress (Kelly 1988, Davis and Kelly 1997); thus, this factor is significant in the positive direction.

	coefficient	standard error	z-value	p-value
intercept	−2.333	.059	−39.5	$< .00001$
initiality	3.787	.062	60.9	$< .00001$
coda size	.723	.046	15.6	$< .00001$

Figure 130: Coda complexity in English nominals.

This model reveals that coda size is aggregately correlated with stress, but does not indicate how specific coda sizes fare with respect to each other. For example, a hierarchy such as C < Ø < CCC < CC (all contrasts significant) could be consistent with this result. To investigate the effect at a finer grain of detail, the coda size factor is separated into four binary factors, one for each level, in figure 131. In this

[85]Another way to set up the model would be to treat the whole disyllable as the datum, rendering the independent variable whether the disyllable is initially or finally stressed (say, 1 and 0, respectively). Then, the coda sizes of both syllables could be entered as separate independent variables. The present model was chosen because it is simpler, with only one coda size effect (cf. figure 137 for treatment initial and final syllables separately). A potential concern with the present formulation is that the stresses of the two syllables of a disyllable are not independent; they are mutually exclusive (for primary stress). To correct for this dependence, one might (as a rule of thumb) double the given p-values. Doing so does not qualitatively alter any of the findings of this section. Moreover, the p-values given here are arguably already overconservative, in that they are two-tailed when (smaller) one-tailed values would arguably suffice (fn. 82).

regression table, the coda factors are forward-difference coded, meaning that each is to be interpreted with respect to the previous factor in the table, as indicated by the COMPARANDUM column (see §2.4.1). Coding the factors in this manner reveals that the effect of coda size in nominals is monotonic increasing: Ø < C < CC < CCC (every contrast significant at $p < .005$). Such a hierarchy is sensible in light of the generalization that increasing complexity (in timing slots of any kind) is expected to correlate with (if anything) greater weight in a scalar weight system (see §7). Under the hypothesis that weight in English is exclusively binary, on the other hand, a contrast such as CC < CCC (which is as great in magnitude as the other contrasts) is unmotivated.

	comparandum	coefficient	standard error	z-value	p-value
intercept	[i.e. Ø]	−1.207	.116	−10.4	< .00001
C	[vs. Ø]	.875	.065	13.4	< .00001
CC	[vs. C]	.410	.095	4.3	= .00002
CCC	[vs. CC]	1.344	.453	3.0	= .003
initiality		3.827	.065	59.2	< .00001

Figure 131: Coda complexity in English nominals (forward-difference coded).

The same generalization emerges independently among the verbals. Figure 132 is the regression table, organized like figure 131 but now based on the verbal rather than nominal subset of data. The initiality factor has the opposite sign for the verbals, indicating that stress is usually final in verbals, unlike nominals (a difference known to be productive; Kelly 1988, Davis and Kelly 1997, Guion et al. 2003). The hierarchy for coda complexity is Ø < C < {CC, CCC} (both significant contrasts $p < .0001$). CC and CCC are not significantly different from each other, but note that CCC codas are rare in verbal roots, being attested in only nine syllables in the present corpus. With this few data points, a nonsignificant result is hardly telling.

	comparandum	coefficient	standard error	z-value	p-value
intercept	[i.e. Ø]	1.258	.272	4.6	$< .00001$
C	[vs. Ø]	.665	.074	9.0	$< .00001$
CC	[vs. C]	1.660	.163	10.2	$< .00001$
CCC	[vs. CC]	$-.046$	1.079	$-.0$	$= .97$
initiality		$-.931$	.073	-12.7	$< .00001$

Figure 132: Coda complexity in English verbals.

In addition to coda complexity, vowel length is correlated with primary stress placement, such that V̆ < VV ($p < .0001$, again independently in both the nominals and verbals); see also Guion et al. (2003) for support of this contrast. For this test, I tentatively bifurcate the English vowels as follows: short = {ə, ɚ, ɪ, ɛ, ʊ, ʌ, æ, ɔ}; long = {a͡ɪ, a͡ʊ, o͡ɪ, e͡ɪ, o͡ʊ, iː, uː, ɑ, ɹ̩}. The result remains significant with various other reasonable length assignments or exclusions. For example, it remains significant if the vowels {ɑ, ɔ, æ} are put aside (as ambiguously long). It also remains significant if the reduced vowels [ə] and [ɚ] are excluded (with or without [ɪ], which often represents a reduced vowel in these data); all the other vowels can be either primary stressed or not (recall that secondary stress and no stress are conflated here).

When the skeletal structure of the rime as a whole is considered, the same hierarchy found consistently in the case studies on quantitative meters in part I, namely, V̆ < V̆C < VV < VVC, is also found here for English stress (with every contrast $p < .001$). In fact, this hierarchy is attested independently in both the nominal and verbal subsets. This parallelism between the parts of speech is illustrated in figure 133. Every syllable in the data is coded for its skeletal structure, including onset. The percentage of the time that each skeletal type is primary stressed is computed separately for nominals and verbals; these are the x- and y-axes of the plot. The size of each labeled point is proportional to its log frequency in the data. The four largest (most frequent) syllable types are precisely the four rime types just mentioned, but with simple onsets: CV̆ < CV̆C < CVV < CVVC. These four types closely approximate the regression line, which is based on all the data (not just these four types).

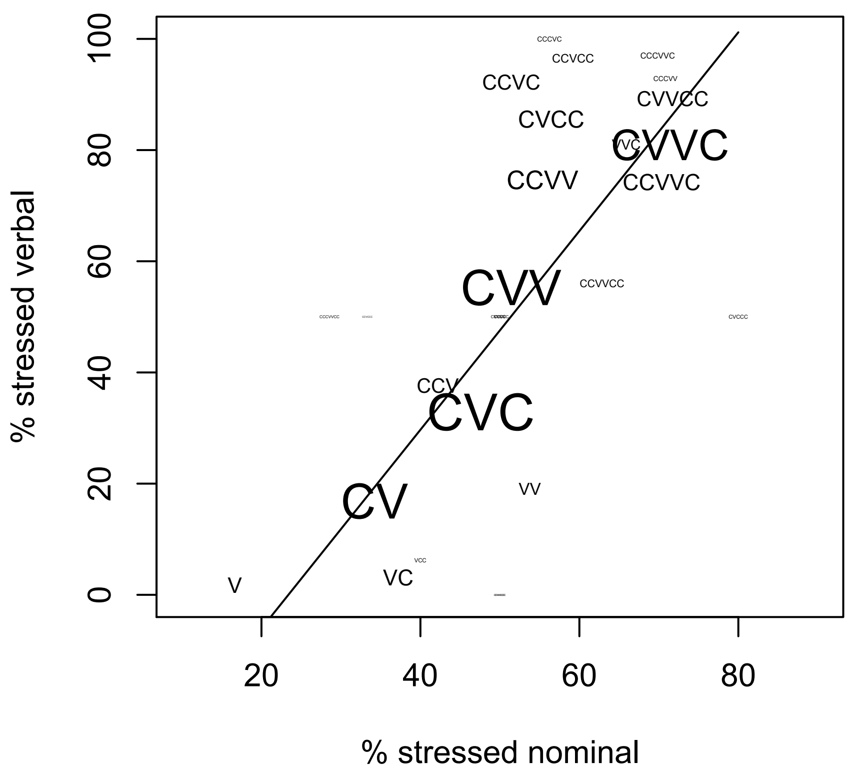

Figure 133: The parallel treatment of syllable structure and weight in nominals and verbals.

A composite hierarchy for nominals and verbals is given in figure 134, weighting each category according to its frequency in the data (thus, nominals are given more weight than verbals). Only the five most frequent rime types in English are shown. Their relative differences on the continuum are to scale (e.g. the V̆C < VV contrast is over twice as great in magnitude as the VV < V̆CC contrast). Thus, as with the studies of weight in quantitative meters, it is insufficient to characterize weight in English stress as being exclusively a hierarchy of weight grades; differences between pairs of types are quantifiable, at least under the present diagnostic.

Figure 134: Stress-attractingness of the five most frequent English rime types.

15.2 Onset structure

Onset length also contributes to the likelihood that an English syllable receives main stress. In fact, as with codas, the effect is monotonic, and found independently in both nominals and verbals. Specifically, among nominals, the hierarchy is Ø < C < CC (<) CCC, with all but the final contrast being highly significant. Among verbals, the same result obtains with the same caveat: Ø < C < CC (<) CCC. See figures 135 and 136 for the specific findings.

	comparandum	coefficient	standard error	z-value	p-value
intercept	[i.e. Ø]	−3.442	.104	−33.1	$< .00001$
C	[vs. Ø]	.914	.097	9.5	$< .00001$
CC	[vs. C]	.313	.079	4.0	$= .00007$
CCC	[vs. CC]	.511	.267	1.9	$= .056$

Figure 135: Onset complexity in English nominals.

	comparandum	coefficient	standard error	z-value	p-value
intercept	[i.e. Ø]	−1.043	.172	−6.1	$< .00001$
C	[vs. Ø]	2.108	.152	13.9	$< .00001$
CC	[vs. C]	1.203	.124	9.7	$< .00001$
CCC	[vs. CC]	1.164	.552	2.1	$= .035$

Figure 136: Onset complexity in English verbals.

These results corroborate those of Kelly (2004) on onset weight — or at least a correlation between stress attraction and onset complexity — in English. They are also consistent with the findings for quantitative meter in part III in which onset complexity was shown to correlate with gradient metrical weight in several traditions (see §13 above for discussion of the issue).

In short, the same syllabic characteristics that were found to correlate with placement in strong positions in quantitative meters are found in English stress as stress-attracting features, such that (in terms of skeletal structure), rime $\breve{\mathrm{V}} < \breve{\mathrm{V}}\mathrm{C} < \mathrm{VV} <$ VVC, onset $\varnothing < \mathrm{C} < \mathrm{CC}$, etc. These scales also align with scales inferred (in part by transitivity) from the crosslinguistic typology of categorical weight criteria (Gordon 2002, 2006). For example, though some languages (e.g. Kashmiri in §14) exhibit the three-way hierarchy $\breve{\mathrm{V}} < \breve{\mathrm{V}}\mathrm{C} < \mathrm{VV}$ in their stress systems, the same hierarchy can be inferred by transitivity from the prevalence of $\breve{\mathrm{V}} < \breve{\mathrm{V}}\mathrm{C}$ and $\breve{\mathrm{V}}\mathrm{C} < \mathrm{VV}$ binary hierarchies (and the virtual absence of their reversals).

One might wonder how many of the aforementioned contrasts might be attributed to rules or conventions other than weight under more traditionally-oriented analyses of English stress. For example, isn't it possible that weight per se is binary, but rimal $\breve{\mathrm{V}}\mathrm{C} < \mathrm{VV}$, for one, is an artifact of final C extrametricality? Under this view, $\breve{\mathrm{V}}\mathrm{C}$ and VV would both be equally heavy non-finally, but finally, $\breve{\mathrm{V}}\mathrm{C}$ would be equivalent to $\breve{\mathrm{V}}$, i.e., light.

Even insofar as the hierarchy could be reanalyzed according to such rules or conventions, one must be wary that the resulting system would conspire to create to surface pattern that accidentally (on such a view) recapitulates the universal weight hierarchy. In other words, if $\breve{\mathrm{V}}\mathrm{C} < \mathrm{VV} < \mathrm{VVC}$ is taken not to be a fact about weight in English, then one must regard its neat alignment with the weight typology (e.g. uncontroversial cases of $\breve{\mathrm{V}}\mathrm{C} < \mathrm{VV}$ and $\mathrm{VV} < \mathrm{VVC}$ as weight in various languages) as coincidental.

Moreover, as an empirical matter, $\breve{\mathrm{V}}\mathrm{C} < \mathrm{VV}$ cannot be written off to final extrametricality: It is found in both initial and final syllables independently. Indeed, as figure 137 summarizes, every hierarchy discussed in this section is found independently in the initial and final syllables of the disyllables (now pooling nominals and verbals into a single data set). There is therefore striking consistency not only across parts of speech, but also across positions in the word, at least in disyllables, reinforcing the universality of the scales both within and across languages.

	initial syllable	final syllable
coda	$\emptyset < C < CC(C)$	$\emptyset < C < CC < CCC$
onset	$\emptyset < C < CC < CCC$	$\emptyset, C < CC(C)$
vowel	$V < VV$	$V < VV$

Figure 137: Comparable complexity effects in disyllable-initial and -final positions.

Of course, highlighting these hierarchies is not to imply that stress assignment in English disyllables is based solely on weight or that every aspect of the system is natural. The purpose is only to demonstrate that, from a broad, data-driven view, the statistical tendencies in English stress, even when they fall short of categoricity, align with weight universals.

16 The productivity of gradient weight in English

16.1 Experimental evidence

In §15, effects of skeletal structure on weight were inferred from the distribution of main stress vis-à-vis syllable structure in a corpus of extant disyllabic words of English, controlling for possible confounds such as initiality, part of speech, and morphological complexity. In this section, I provide some direct evidence that these effects are a productive aspect of the English grammar of stress assignment, being extended to novel (also known as NONCE or WUG) words.

To gauge productivity experimentally, I conduct a wug test using Amazon's Mechanical Turk (`www.mturk.com`), an online platform connecting laborers to *ad hoc* tasks (see Daland et al. 2011:203 on some virtues of this approach). Participants are screened in several ways. First, only users with a prior approval rating of 95% or higher are permitted to access the task (*ibid.*); another prerequisite is location in the United States according to Amazon's records. Participants are also required to give their informed consent and self-report their proficiency in English. Only data from speakers reporting 'high' or 'native' proficiency are used here, though all participants are paid regardless of reported proficiency. Participants also report their geographic dialect of English, from a list of options plus 'other', though this information is not used in screening.

Finally, data are accepted only from participants scoring at least 4/5 on training items (though, once again, all participants who complete the task are paid regardless of their performance on screening criteria). These items are the same for all participants, namely, the nouns *Simon*, *Michael*, *bamboo*, *Kathleen*, and *Arnold* (in that order). These items were selected to exhibit a mixture of initial and final stress and to be words on which nonnative speakers (or native speakers with a poor grasp of the task) might easily err. Participants are asked to indicate whether each word is stressed initially or finally. In both the training and test items, a word (real or nonce) is presented in all caps, followed by a choice between initial vs. final stress (not necessarily in that order), as indicated by capitalizing only one of the two syllables. For example, the first prompt is 'SIMON ○ siMON ○ SImon', where the participant must click exactly one of the two circles. The choices are always presented in pseudorandom order, such that the initially stressed option is first 50% of the time.

The general instructions are reproduced in figure 138. The instructions for the practice items are: ‘Here are 5 actual words/names of English. Indicate whether you pronounce each one with stress on the first or second syllable (as indicated by capitalization, e.g. ENGlish, chiNESE).’

> You will see 25 English words, some real and some made-up (some of which which look rather strange). **Pronounce each to yourself**, and decide whether you place more stress (emphasis) on the first syllable (as in ZEBra) or on the second (as in gaZELLE).
>
> You must select an answer for all words, including practice words, in order to receive payment.
>
> Do not complete multiple versions of this test within the same 12-hour period. If you complete multiple HITs from this HIT group, ONLY THE FIRST ONE WILL BE ACCEPTED.
>
> Your ratings will be used for scientific research, so we must obtain your informed consent before you can begin. In addition, we will ask about your language background, and ask you to practice on 5 real English words.

Figure 138: Wug test instructions.

After completing the training items, each participant is presented with 25 prompts in random order, including 13 wugs and 12 real words (as fillers), the latter being selected at random from the corpus of monomorphemic disyllables in §15. Because all training items and interspersed fillers are nouns, it is intended (though neither crucial nor made explicit) that the wugs also be read as nouns. Each prompt is accompanied by two responses in random order, as above. Participants are paid 24 cents each for completing the task, and not allowed to repeat the task with different test items. The experiment was completed by 400 participants, of whom 220 passed the selection criteria (making for $220 \times 13 = 2860$ wug tokens in the usable data). Average time to complete the experiment was 2.8 minutes, though no time limit was imposed.

Disyllabic wugs are randomly generated according to a fixed construction schema. First, an onset is selected from the following list (all lists given in orthography): {*p*, *b*, *t*, *d*, *f*, *sn*, *sm*, *pr*, *br*, *cr*, *gr*, *pl*, *bl*, *cl*, *gl*, *tw*, *dw*, *sw*, *fr*, *fl*}. Then, a vowel is added: {*i*, *e*, *ee*, *oo*}. Next, a coda is optionally (with a probability of 0.5) appended: {*t*, *p*, *g*, *d*, *b*, *n*, *m*, *l*, *r*}. Next, the onset of the second syllable is selected. If the first syllable lacks a coda, this onset is selected from {*p*, *t*, *d*, *k*, *m*, *n*, *l*, *b*, *f*}. Immediately following a short vowel (orthographic *i* or *e*), the letter is doubled (e.g. *blinnorp* for intended [blɪnɔɹp], not *blinorp*), reflecting a convention of English spelling. If, on the other hand, a coda has been selected for the first syllable, some more phonotactic sensitivity is in order. In this case, the constructor consults the list of real English disyllables, chooses one at random with the coda in question in the first syllable, and uses the following onset from that word in the wug being constructed. This ensures that complex interludes in wugs are distributed like those in real English monomorphemes and also avoids certain other problems, such as the risk of selecting the same consonant as both the coda and onset, in which case the participant would be expected to read a singleton.[86]

Finally, the remainder of the disyllable is randomly selected from the following list: {*orp*, *eln*, *oom*, *olb*, *alt*, *itz*, *oaf*, *een*, *oil*, *eem*, *eeve*, *oke*, *oor*, *arl*, *aft*, *aine*, *arp*, *oon*, *ent*, *ie*, *oe*, *oo*, *oi*, *ay*, *el*, *il*, *ut*, *uk*, *om*, *ak*, *ap*, *if*}. I refer to these endings as COMPLETIONS.[87] This schema is designed to generate a relatively open-ended set of disyllables that are relatively unambiguous in terms of pronunciation, given the

---

[86]In syllabifying the disyllables, onset maximization is assumed (see §6), such that a word such as *abrupt* would not count as a coda-*b* word but *abduct* would. It follows that the constructor would never recruit a sonorant as an onset following a stop coda, since such the intended coda would presumably be syllabified as part of an onset in such a context.

[87]A few of these completions resemble extant suffixes. This is not a concern, however, as the influence of the completion is factored out (see below).

conventions of English orthography. Furthermore, the final syllables tend to be on the heavier side in order to compensate somewhat for the overwhelming tendency for English nouns to receive initial stress (§15). Figure 139 exemplifies 12 wugs generated according to this procedure.

blertaine	crinnorp
glennolb	deekoor
fleebsoom	smebyeln
crekkoe	smirsap
grenneln	sweemarl
plibboaf	dendaft

Figure 139: Sample wug disyllables.

16.2 Results and discussion

This section summarizes the findings for effects of skeletal structure on stress propensity in wugs. Only the initial syllables of the wugs are analyzed. The smaller set of completions (i.e. word-final VC_0 strings) are modeled as random effects.[88] The data therefore include 2,860 syllable tokens (one for each wug). Given the constructor in §16.1, the rime of each token can only be $\breve{V}$, $\breve{V}$C, VV, or VVC. The effect of interest here is how these different rime shapes influence the dependent variable, that is, whether or not the syllable receives main stress. Because this outcome is binary (1 = stressed, 0 = unstressed), logistic regression is used. In figure 140, as usual, the rime structure factors are forward-difference coded.

Two additional factors are included as fixed effects in the model summarized in figure 140, though both are nonsignificant. First, `prevIsStressed` is 1 iff the

[88] Thus, any differences in stress-attractingness among completions are effectively controlled. At any rate, the design ensures that the initial syllable and completion are independent of each other, so one would not expect the shape of the completion to be a confound.

participant chose initial stress for the previous prompt. If this effect were significant, it might suggest some kind of priming or habituation across prompts. Second, `isFirstAns` is 1 iff the chosen response is the leftmost of the pair, counteracting any bias participants might have in the aggregate for initial (leftmost) vs. final (rightmost) responses. Recall that whether the initially-stressed answer is presented first or second is random for each prompt. The model includes two random effects (not shown in the figure), first, the completion (e.g. *orp*, *ap*), and second, the participant's numerical identification code (effectively assigning a baseline preference for initial stress for each participant).[89]

	comparandum	coefficient	standard error	z-value	p-value
intercept	[i.e. V̆]	1.125	.148	7.6	$< .00001$
V̆C	[vs. V̆]	.727	.122	6.0	$< .00001$
VV	[vs. V̆C]	.397	.133	3.0	$= .003$
VVC	[vs. VV]	.484	.141	3.4	$= .0006$
prevIsStressed		$-.042$	.105	$-.4$	$= .69$
isFirstAns		$-.087$	.092	$-.9$	$= .34$

Figure 140: Regression table for rime shape in wugs.

The rimal hierarchy V̆ < V̆C < VV < VVC (all contrasts $p < .005$) is therefore observed not only in the lexicon in English but also in stress assignment in wugs, as shown here, supporting its productivity. Note that the contrast V̆C < VV cannot be attributed to final extrametricality, given that only nonfinal syllables are considered. Moreover, it cannot be attributed a conceivable confound from orthography: Given the experimental methodology, one might entertain a possible counteranalysis of these results in which participants are more likely to choose the answer with the initial

[89] A different formulation of the model was also tested in which `prevIsStressed` and `isFirstAns` are both conditioned on participant instead of fixed (global) effects. The reported findings are not qualitatively altered by this adjustment.

syllable capitalized if the initial syllable contains more letters, irrespective of its phonological weight. However, given the constructor in §16.1, $\breve{\mathrm{V}}$C and VV both always comprise exactly two letters (compare, e.g., *BEEnorp*, with a VV rime, and *BINdorp*, with a $\breve{\mathrm{V}}$C rime); this confound is therefore not a concern for this contrast. It is also not an issue for $\breve{\mathrm{V}} < \breve{\mathrm{V}}$C, since $\breve{\mathrm{V}}$ syllables likewise always have exactly two letters in the orthographic rime, given the spelling and capitalization conventions assumed here (compare, e.g., *BINnorp*, with a $\breve{\mathrm{V}}$ rime, and *BINdorp*, with a $\breve{\mathrm{V}}$C rime). This orthographic confound is potentially relevant, however, for the contrast VV < VVC, as VVC usually has one more letter than VV in the orthographic syllable. (Still, VV < VVC was also found in the lexicon in §15, where this confound is moot.)

Guion et al. (2003) likewise find that rime $\breve{\mathrm{V}}$ < VV (as well as $\breve{\mathrm{V}}$CC < VVC) using a different methodology to wug-test English disyllables. For example, in one test, they auditorily present participants with two stressed syllables with an intervening pause. Participants are asked to concatenate the two syllables into a single word, which they then pronounce aloud in a frame. While the results of Guion et al. (2003) are consistent with the present findings, they did not specifically investigate any of the binary contrasts $\breve{\mathrm{V}} < \breve{\mathrm{V}}$C, $\breve{\mathrm{V}}$C < VV, or VV < VVC, as is done here. See also Shelton (2007), who concludes from various experiments on the role of rising vs. falling diphthongs in Spanish stress that it is insufficient to regard weight as binary in that language.

The effect of onset complexity on stress placement described for the lexicon in §15 is also corroborated by these experimental data. The initial syllables of the wugs used here comprise either one to two consonants. When a fixed effect for onset complexity is added to the model in figure 140, coded as a binary factor, it is significant (p = .0001) and positive (coefficient = .29) — not as strong an effect as the rime contrasts, but also not far off in magnitude from $\breve{\mathrm{V}}$C < VV, the weakest of the three rime contrasts. The correlation of onset complexity and stress once again corroborates Kelly (2004). Nevertheless, both Kelly (2004) and the present experiment, being based on written prompts, are subject to a possible confound from orthography. In particular, one might consider whether the experimental participants tend to assign stress to visually or orthographically larger syllables, irrespective of phonology. Given that the effect of onset complexity is robustly attested (even in mutually independent tests) in the lexicon (§15), it seems unlikely that this orthographic confound could be the entire story, but it is a possible confound for the finding that onset C < CC in the present experiment.[90]

16.3 Local summary

In §15 and §16, various aspects of the skeletal structure of the syllable are shown to significantly correlate with the likelihood that a syllable receives primary stress in its disyllable, both in the lexicon and in experimental nonce probes. For example, in both domains, the following hierarchy of stress attraction is observed for rimes: $\breve{V}$ < $\breve{V}$C < VV < VVC. This is the same hierarchy found for quantitative meters in part I.[91] It is also demonstrated here (corroborating Kelly 2004) that onset complexity is positively correlated with stress both in the lexicon and in wugs. In fact, both the coda and onset correlations are monotonic increasing, such that Ø < C < CC (<) CCC (at least in the lexicon; not all of these contrasts were tested in wugs). This same correlation was found for onset sensitivity in quantitative meters (§13).

If one defines weight functionally — (progressively) heavier syllables are (pro-

[90]This confound be controlled for by including digraphs (e.g. *sh*, *ph*) in the data and comparing them to complex onsets, holding orthographic complexity constant.

[91]See §7 on the general principles of weight motivating it, including complexity and sonority.

gressively) greater attractors of stress (cf. §1 on defining weight in metrics) — these continua can be regarded as reflecting scalar weight. This move then explains why they all abide by weight universals, aligning neatly with other weight-sensitive phenomena crosslinguistically: They are weight. If, on the other hand, one regards weight in English stress as being exclusively binary, such that the majority of effects discussed here are due to factors other than weight, one would have to find a constellation of independent motivations for the contrasts documented here. But even then, one would still be left with an apparent conspiracy: The various non-weight factors would add up to a system in which every one of the numerous contrasts in stress attraction (see figure 137 for 11 examples) happens to abide by weight universals. For example, if the $\breve{V}C < VV$ contrast (both in final and non-final syllables) is not a fact about weight, then it is an accident that it aligns with the near-universal polarity of this contrast in systems that are uncontroversially weight.[92] Regarding these contrasts as weight is a restrictive hypothesis, in that one predicts that they will strongly tend to be consistent with weight universals.

17 Conclusion

Canonical weight-sensitive systems in phonology, including quantitative meter and quantity-sensitive stress, are proposed to sometimes be sensitive to a gradient continuum of syllable weight. In such (sub)systems, weight is properly characterized not as a strict domination hierarchy (i.e. ordinal scale) of categories, but as an interval scale in which (1) there is no clear segregation of syllable types into categories and (2) weight differences between syllable types are statistically quantifiable. The role of

[92] Cf. §10.2 for arguments against a variable moraicity analysis of $\breve{V}C < VV$.

categoricity in weight is not denied here; rather, categoricity and gradience are argued to coexist (to greater or lesser extents) in weight-sensitive systems. This is accomplished through a constraint-based framework in which categorically and gradiently evaluated constraints interact to generate the weight mapping typology (§10).

This typology includes purely categorical systems (such as Yana stress in figure 128), purely (or close to purely) gradient systems (Kamban̠'s Tamil meter being a possible case; §9), and mixed categorical-gradient systems, in which a binary criterion is clearly in force, but additional intra-categorial sensitivity to weight is also observed (see, e.g., the analyses of weight mapping in the Greek hexameter and the Kalevala meter in §10.1). These mixed systems provide evidence for incomplete categorization (or incomplete phonologization), in which a phonological system exhibits polarization towards categories, but is not entirely categorical, in the sense that the relevant phonetic interface continues to leak through within those categories.

Finally, analyzing gradient weight systems strengthens and extends the evidence for weight universals. Consider, for instance, the relevance of onset structure to syllable weight, particularly the distinction between a null (Ø) and simple (C) onset. If this distinction is relevant for assigning syllables to weight categories (as in Pirahã; §14), a null onset universally patterns as lighter than a simple onset: Ø < C (Gordon 2005, Topintzi 2010; §13). In most languages, however, weight categorization is blind to onset structure. As I demonstrate here, even in such languages (e.g. English and Sanskrit), onset structure exerts a statistically significant effect on weight in non-categorical contexts, reflecting sensitivity to the same distinctions that rise to the level of categoricity in other languages, such as Pirahã. In almost every instance, these gradient contrasts (e.g. onset Ø < C) align in polarity with their categorical counterparts in other languages. Put succinctly, individual languages are like microcosms of the crosslinguistic typology in the gradient realm.

REFERENCES

Allen, W. Sidney. 1973. *Accent and rhythm*. London, U.K.: Cambridge University Press.

Allen, W. Sidney. 1978. *Vox Latina: A guide to the pronunciation of Classical Latin*. London, U.K.: Cambridge University Press.

Allen, W. Sidney. 1983. Some reflections on the penultimate accent. *Illinois Classical Studies* 8.

Árnason, Kristján. 1991. *The rhythms of dróttkvætt and other Old Icelandic metres*. Reykjavík: Institute of Linguistics, University of Iceland.

Árnason, Kristján. 1998. Review of K. E. Gade, the structure of Old Norse Dróttkvætt poetry, Islandica 49, Ithaca: Cornell University Press, 1995. *Alvíssmál* 8.98–109.

Árnason, Kristján. 2009. On Kuhn's laws and Craigie's law in Old Icelandic poetry. *Versatility in versification: multidisciplinary approaches to metrics*, ed. by Tonya Kim Dewey, 39–59. Peter Lang.

Arnold, Edward Vernon. 1905. *Vedic metre in its historical development*. Cambridge University Press.

Asher, R. E., and T. C. Kumari. 1997. *Malayalam*. London, U.K.: Routledge.

Aujac, Germaine, and Maurice Lebel. 1981. *Denys d'Halicarnasse: Opuscules rhétoriques. Tome 3. La composition stylistique*. Paris: Les Belles Lettres.

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59.390–412.

Baayen, R. H., R. Piepenbrock, and L. Gulikers. 1995. The CELEX lexical database (release 2) [CD-ROM]. Philadelphia, PA: Linguistics Data Consortium, University of Pennsylvania.

Baayen, R. Harald. 2004. Statistics in psycholinguistics: a critique of some current gold standards. *Mental Lexicon Working Papers* 1.1–47.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statis-*

tics using R. Cambridge, U.K.: Cambridge University Press.

Bates, Douglas, and Martin Maechler. 2009. *lme4: Linear mixed-effects models using S4 classes*. URL `CRAN.R-project.org/package=lme4`, R package version 0.999375-31.

Bauer, Winifred. 1993. *Maori*. New York: Routledge.

Beckman, Jill N. 1998. *Positional faithfulness*. Doctoral Dissertation, University of Massachusetts, Amherst.

Beckman, Nataniel. 1899. Kritische Beiträge zur altnordischen Metrik. *Arkiv för nordisk filologi* 15.

Bennett, Charles E. 1918. *New Latin grammar*. Ithaca, New York: Allyn and Bacon.

Benor, Sarah, and Roger Levy. 2006. The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82.233–278.

Berko, J. 1958. The child's learning of English morphology. *Word* 14.150–177.

Biggs, Henry. 1996. *A statistical analysis of the metrics of the classic French decasyllable and classic alexandrine*. Doctoral Dissertation, University of California, Los Angeles.

Blevins, Juliette. 2003. The phonology of Yurok glottalized sonorants: Segmental fission under syllabification. *International Journal of American Linguistics* 69.371–396.

Blevins, Juliette, and Andrew Garrett. 2004. The evolution of metathesis. *Phonetically-based phonology*, ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade, 117–156. Cambridge, U.K.: Cambridge University Press.

Blumenfeld, Lev. 2010. Coercion and minimality. Ms., Carleton University.

Boas, Franz. 1947. Kwakiutl grammar with a glossary of the suffixes. *Transactions of the American Philosophical Society*, ed. by Helene Boas Yampolsky and Zelig Harris, volume 37.3, 201–377.

Boersma, Paul, and Joe Pater. 2008. Convergence properties of a gradual learning algorithm for harmonic grammar. Ms, University of Amsterdam and University of Massachusetts, Amherst.

Boersma, Paul, and David Weenink. 2011. Praat: doing phonetics by computer. Computer program, version 5.2.23, retrieved 1 May 2011 from `http://www.praat.org/`.

Bradley, Travis G. 2001. *The phonetics and phonology of rhotic duration contrast and neutralization*. Doctoral Dissertation, The Pennsylvania State University.

Broselow, Ellen, Susan Chen, and Marie Huffman. 1997. Syllable weight: convergence of phonology and phonetics. *Phonology* 14.47–82.

Brown, Charles P. 1981. *A grammar of the Telugu language*. New Delhi: Asian Educational Services.

Burzio, Luigi. 1994. *Principles of English stress*. Cambridge, U.K.: Cambridge University Press.

Carpenter, Angela C. 2010. A naturalness bias in learning stress. *Phonology* 27.345–392.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. Cambridge, Massachusetts: Massachusetts Institute of Technology Press.

Christdas, Prathima. 1988. *The phonology and morphology of Tamil*. Doctoral Dissertation, Cornell University.

Coltheart, M. 1981. The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology* 33.497–505.

Cooper, William, and John R. Ross. 1975. World order. *Papers from the Parasession on Functionalism*, ed. by R. Grossman, L.J. San, and T. Vance, 63–111. Chicago: Chicago Linguistic Society.

Coulson, Michael. 1992. *Sanskrit: An introduction to the classical language*. Hodder & Stoughton. Revised by Richard Gombrich and James Benson.

Craigie, William. 1900. On some points in skaldic metre. *Arkiv för nordisk filologi* 16.341–384.

Crosswhite, Katherine. 2006. An auditory approach to phonological prominence. Paper presented at the 14th Manchester Phonology Meeting.

Crowhurst, Megan J., and Lev D. Michael. 2005. Iterative footing and prominence-driven stress in Nanti (Kampa). *Language* 81.47–95.

Daland, Robert. 2009. *Diphone-based lexical bootstrapping*. Doctoral Dissertation, Northwestern University.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis, and Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28.197–234.

Davis, S., and M. Kelly. 1997. Knowledge of the English noun-verb stress difference by native and nonnative speakers. *Journal of Memory and Language* 36.445–460.

Davis, Stuart. 1988. Syllable onsets as a factor in stress rules. *Phonology* 5.1–19.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum Likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39.1–38.

Deo, Ashwini. 2007. The metrical organization of Classical Sanskrit verse. *Journal of Linguistics* 43.

Devine, Andrew M., and Laurence Stephens. 1975. The abstractness of metrical patterns: Generative metrics and traditional explicit metrics. *Poetics* 16.411–430.

Devine, Andrew M., and Laurence Stephens. 1976. The Homeric Hexameter and a basic principle of metrical theory. *Classical Philology* 71.141–163.

Devine, Andrew M., and Laurence Stephens. 1977. Preliminaries to an explicit theory of Greek metre. *Transactions of the American Philological Association* 107.103–129.

Devine, Andrew M., and Laurence Stephens. 1994. *The prosody of Greek speech*. Oxford University Press.

DiCanio, Christian. Forthcoming. Cross-linguistic perception of Itunyoso Trique tone. MS, CNRS / Laboratoire Dynamique du Langage, Université Lyon 2.

Duckworth, G. E. 1969. *Vergil and classical hexameter poetry*. Ann Arbor, Michigan: University of Michigan Press.

Dupoux, E., Y. Hirose, K. Kakehi, C. Pallier, and J. Mehler. 1998. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25.1568–1578.

Efron, B., and R. J. Tibshirani. 1994. *An introduction to the bootstrap*. Boca Raton, FL: CRC Press.

Efron, Bradley. 1979. Bootstrap methods: another look at the Jackknife. *The Annals of Statistics* 7.1–26.

Emeneau, Murray B. 1939. The vowels of the Badaga language. *Language* 15.43–47.

Everett, Daniel. 1988. On metrical constituent structure in Pirahã. *Natural Language and Linguistic Theory* 6.207–246.

Everett, Daniel, and Keren Everett. 1984. On the relevance of syllable onsets to stress placement. *Linguistic Inquiry* 15.705–711.

Fabb, Nigel, and Morris Halle. 2008. *Meter in poetry. With a chapter on Southern Romance meters by Carlos Piera*. Cambridge, U.K.: Cambridge University Press.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18.

Flemming, Edward. 2003. The relationship between coronal place and vowel backness. *Phonology* 20.335–373.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. *Phonetically based phonology*, ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade, 232–276. Cambridge, U.K.: Cambridge University Press.

Gade, Kari Ellen. 1995. *The structure of Old Norse dróttkvætt poetry*. Ithaca, New York: Cornell University Press.

Gansner, Emden, and Stephen North. 1999. An open graph visualization system and its applications to software engineering. *Software – Practice and Experience* 30.1203–1233.

Garrett, Edward. 1999. Minimal words aren't minimal feet. *Papers in Phonology 2: UCLA Working Papers in Linguistics, Vol. 1*, ed. by Matthew Gordon, 68–105. University of California, Los Angeles.

Getty, Michael. 1998. *A constraint-based approach to the meter of Beowulf*. Doctoral Dissertation, Stanford University.

Goedemans, R. 1998. *Weightless segments*. The Hague: Holland Academic Graphics.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. by Jennifer Spenader, Anders Eriksson, and Osten Dahl, 111–120.

Gordon, Matthew. 1999. *Syllable weight: phonetics, phonology, and typology*. Doctoral Dissertation, University of California, Los Angeles.

Gordon, Matthew. 2002. A phonetically-driven account of syllable weight. *Language* 78.51–80.

Gordon, Matthew. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23.595–653.

Gordon, Matthew. 2006. *Syllable weight: Phonetics, phonology, typology*. Routledge.

Gordon, Matthew, Carmen Jany, Carlos Nash, and Nobutaka Takara. 2008. Vowel and consonant sonority and coda weight: A cross-linguistic study. *Proceedings of the 26th West Coast Conference on Formal Linguistics (WCCFL)*, ed. by Charles B. Chang and Hannah J. Haynie, 208–216. Somerville, MA: Cascadilla Proceedings Project.

Greenough, J. B. 1900. *Bucolics, Aeneid, and Georgics of Vergil*. Boston, Massachusetts: Ginn & Col.

Guion, Susan G., J. J. Clark, Tetsuo Harada, and Ratree P. Wayland. 2003. Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech* 46.403–427.

Halle, Morris. 1970. On meter and prosody. *Progress in linguistics*, ed. by Manfred Bierwisch and Karl Erich Heidolph, 64–80. The Hague: Mouton & Co.

Halle, Morris, and Samuel Jay Keyser. 1971. *English stress: Its form, its growth, and its role in verse*. New York: Harper and Row.

Halle, Morris, and Jean-Roger Vergnaud. 1987. Stress and the cycle. *Linguistic*

Inquiry 18.45–84.

Hanson, Kristin, and Paul Kiparsky. 1996. A parametric theory of poetic meter. *Language* 72.287–335.

Harlow, R. 2001. *A Maori reference grammar*. Auckland, New Zealand: Longman.

Harris, John. 1994. *English sound structure*. Oxford: Blackwell.

Hart, George, and Hank Heifetz. 1988. *The Forest Book of the Rāmāyaṇa of Kampaṉ*. Berkeley, California: University of California Press.

Hart, Kausalya. 1999. *Tamil for beginners*. University of California, Berkeley: International and Areal Studies.

Hayes, Bruce. 1979. The rhythmic structure of Persian verse. *Edebiyāt* 4.193–242.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13.227–276.

Hayes, Bruce. 1988. Metrics and phonological theory. *Linguistics: The Cambridge survey, vol. ii, Lingustic theory: Extensions and implications*, ed. by Frederick J. Newmeyer, volume 2, chapter 12, 220–249. Cambridge University Press.

Hayes, Bruce. 1989a. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20.253–306.

Hayes, Bruce. 1989b. The prosodic hierarchy in meter. *Rhythm and meter: Phonetics and phonology, volume I*, ed. by Paul Kiparsky and Gilbert Youmans. San Diego, CA: Academic Press.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Hayes, Bruce. 1997. Phonetically-driven phonology: the role of Optimality Theory and inductive grounding. *Proceedings of the 1996 Milwaukee Conference on Formalism and Functionalism in Linguistics*.

Hayes, Bruce, and Claire Moore-Cantwell. 2011. Gerard Manley Hopkins's Sprung Rhythm: Corpus study and stochastic grammar. *Phonology* .

Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.379–440.

Hermann, Eduard. 1923. *Silbenbildung im Griechischen und in den anderen indogermanischen Sprachen*. Göttingen: Vandenhoeck & Ruprecht.

Hoenigswald, Henry M. 1990. Overlong syllables in Rigvedic cadences. *Journal of the American Oriental Society* 109.559–563.

Hoenigswald, Henry M. 1991. The prosody of the epic adonius and its prehistory. *Illinois Classical Studies* XVI.1–15.

Hoenigswald, Henry M. 1994. Meter and phonology: The chronological interpretation of idealized reconstructions. *Früh-, Mittel-, Spätindogermanisch: Akten der IX. Fachtagung der Indogermanischen Gesellschaft*, ed. by George E. Dunkel, Gisela Meyer, Salvatore Scarlata, and Christian Seidl, 135–148. Wiesbaden.

Hoequist, Charles E. 1985. *A comparative study of linguistic rhythm*. Doctoral Dissertation, Yale University.

Hoffory, J. 1889. *Eddastudien: erster Theil*. G. Reimer.

Hogg, Richard, and Christopher B. McCully. 1987. *Metrical phonology: a course book*. Cambridge, U.K.: Cambridge University Press.

Holt, Brigitte, and Robert A. Benfer. 2000. Estimating missing data: An iterative regression approach. *Journal of Human Evolution* 39.289–298.

Holtsmark, Erling B. 2010. Enchiridion of metrics. URL `classics-laohu.home.mchsi.com/greek-metrics/`.

Hubbard, Kathleen. 1994. *Duration in moraic theory*. Doctoral Dissertation, University of California, Berkeley.

Hunter, D. R., and K. Lange. 2004. A tutorial on MM algorithms. *The American Statistician* 58.30–37.

Hyman, Larry. 1985. *A theory of phonological weight*. Dordrecht: Foris.

International Phonetic Association. 1999. *Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet*. Cambridge, U.K.: Cambridge University Press.

Introduction to SAS. 2011. UCLA: Academic Technology Services, Statistical Consulting Group. URL `www.ats.ucla.edu/stat/sas/notes2/`.

Irigoin, J. 1965. Review of L. Rossi, Metrica e critica stilistica. *Göttingische Gelehrte Anzeigen* 217.284.

Jacobs, Haike. 1997. Latin enclitic stress revisited. *Linguistic Inquiry* 28.648–661.

Jaeger, T. F. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59.434–446.

Jäger, Gerhard. 2007. Maximum Entropy models and Stochastic Optimality Theory. *Architectures, rules, and preferences: A Festschrift for Joan Bresnan*, ed. by J. Grimshaw, J. Malin, C. Manning, J. Simpson, and A. Zaenen, 467–479. CSLI: Stanford.

Jakobson, Roman. 1933. Über den Versbau der serbokroatischen Volksepen. *Archives néerlandaises de phonétique expérimentale* 7–9.44–53. Reprinted in Jakobson (1966).

Jakobson, Roman. 1966. *Selected writings IV: Slavic epic studies*. The Hague: Mouton.

Johnson, Mark. 2002. Optimality-Theoretic Lexical Functional Grammar. *The lexical basis of sentence processing: Formal, computational, and experimental issues*, ed. by Paula Merlo and Susan Stevenson, 59–74. John Benjamins, Amsterdam.

Kager, René. 1989. *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.

Kamban̲. 1956. *Kampar Iyar̲r̲iya Irāmāyaṇam*. Annamalai, India: Annamalai University Press. Critical edition of the *Kamparāmāyaṇam* (c. 1200 CE).

Kapatsinski, Vsevolod. 2009. Testing theories of linguistic constituency with configural learning: The case of the English syllable. *Language* 85.248–277.

Kawahara, Shigeto. 2011. Experimental approaches in theoretical phonology. *The blackwell companion to phonology*, ed. by M. van Oostendorp, C. Ewen, E. Hume, and K. Rice, 2283–2303. Wiley-Blackwell Publishers.

Keane, Elinor. 2003. Word-level prominence in Tamil. *Proceedings of the 15th Inter-*

national Congress of the Phonetic Sciences, volume 2, 1257–1260.

Keane, Elinor. 2006. Prominence in Tamil. *Journal of the International Phonetic Association* 36.1–20.

Keating, P., T. Cho, C. Fougeron, and C. Hsu. 2003. Domain-initial articulatory strengthening in four languages. *Phonetic interpretation*, ed. by J. Local, R. Ogden, and R. Temple, volume 6 of *Papers in Laboratory Phonology*, 143–161. Cambridge, U.K.: Cambridge University Press.

Kelly, Michael. 1988. Phonological biases in grammatical category shifts. *Journal of Memory and Language* 27.343–358.

Kelly, Michael. 2004. Word onset patterns and lexical stress in English. *Journal of Memory and Language* 50.231–244.

Kessler, Brett. 1998. Ambisyllabicity in the language of the Rigveda. MS, Stanford University.

Kessler, Brett. 2001. *The significance of word lists*. Stanford, California: CSLI Publications.

Kessler, Brett, and Rebecca Treiman. 1997. Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language* 37.295–311.

Ketner, Katherine Heidel. 2006. *Size restrictions in Prosodic Morphology*. Doctoral Dissertation, University of Cambridge.

Keyser, Samuel Jay, and Paul Kiparsky. 1984. Syllable structure in Finnish phonology. *Language sound structure*, ed. by Mark Aronoff and Richard Oehrle. Cambridge, Massachusetts: Massachusetts Institute of Technology Press.

Kiparsky, Paul. 1968. Metrics and morphophonemics in the Kalevala. *Studies presented to Roman Jakobson by his students*, ed. by Charles Gribble. Cambridge, Massachusetts: Slavica.

Kiparsky, Paul. 1972. Metrics and morphophonemics in the Rigveda. *Contributions to generative phonology*, ed. by Michael K. Brame, 171–200. Austin: University of Texas Press.

Kiparsky, Paul. 1977. The rhythmic structure of English verse. *Linguistic Inquiry* 8.189–248.

Kiparsky, Paul. 1989. Sprung rhythm. *Rhythm and meter*, ed. by Paul Kiparsky and Gilbert Youmans. San Diego, CA: Academic Press.

Kiparsky, Paul. 2003. Finnish noun inflection. *Generative approaches to Finnic linguistics*, ed. by Diane Nelson and Satu Manninen. Stanford, California: CSLI Publications.

Kiparsky, Paul. 2010. Compensatory lengthening. To appear in a handbook on the syllable, ed. C. Cairns and E. Raimy, Brill.

Klein, Jared S. 2002. Responsion in the Rigveda. *Journal of the American Oriental Society* 122.311–317.

Krishnamurti, Bhadriraju. 2003. *The Dravidian languages*. Cambridge, U.K.: Cambridge University Press.

Kuhn, Hans. 1983. *Das dróttkvætt*. Heidelberg: Winter.

de Lacy, Paul. 1997. Prosodic categorization. Available on Rutgers Optimality Archive, ROA-236, `http://ruccs.rutgers.edu/roa.html`.

de Lacy, Paul. 2002. *The formal expression of markedness*. Doctoral Dissertation, University of Massachusetts-Amherst.

de Lacy, Paul. 2004. Markedness conflation in Optimality Theory. *Phonology* 21.145–199.

Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*. Blackwell Publishing.

Lauerma, Petri. 2001. Larin Parasken metriikasta. *Virittäjä* 1.

Lehmann, Thomas. 1994. *Grammatik des Alttamil unter besonderer Berücksichtigung der Cankam-Texte des Dichters Kapilar*. Stuttgart: Franz Steiner Verlag. Revision of 1992 Ph.D. dissertation, University of Heidelberg.

Leino, Pentti. 1986. *Language and metre. Metrics and the metrical system of Finnish*. Number 31 Studia Fennica. Helsinki: Suomalaisen Kirjallisuuden Seura.

Leino, Pentti. 1994. The Kalevala Metre and its development. *Songs beyond the Kalevala: Transformations of oral poetry*, ed. by Anna-Leena Siikala and Sinikka Vakimo, 56–74. Helsinki: SKS.

Leskinen, Heikki, and Jaakko Lehtonen. 1985. Zur wortphonologischen Quantität in den Südostdialekten des finnischen. *Studia Fennica* 28.49–83.

Levy, Roger. 2010. Probabilistic models in the study of language. MS for a textbook, UC San Diego.

Lewin-Koh, Nicholas J., and Roger Bivand. 2010. *Maptools: tools for reading and handling spatial objects*. URL `http://CRAN.R-project.org/package=maptools`, R package version 0.7-38.

Lin, Ying. 2005. *Learning features and segments from waveforms: A statistical model of early phonological acquisition*. Doctoral Dissertation, University of California, Los Angeles.

Lönnrot, Elias. 1849. *Kalevala taikka vanhoja Karjalan runoja Suomen kansan muinoisista ajoista*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Lunden, Anya. 2006. *Weight, final lengthening and stress: A phonetic and phonological case study of Norwegian*. Doctoral Dissertation, UC Santa Cruz.

Maas, Paul. 1962. *Greek metre*. Oxford University Press.

Macdonell, Arthur Anthony. 1916. *A Vedic grammar for students*. Oxford University Press.

Macdonell, Arthur Anthony. 1927. *A Sanskrit grammar for students*. Oxford University Press.

Maddieson, Ian. 1993. Splitting the mora. *UCLA Working Papers in Phonetics* 83.9–18.

Manaster Ramer, Alexis. 1981. *How abstruse is phonology?*. Doctoral Dissertation, University of Chicago.

Manaster Ramer, Alexis. 1994. Stefan George and phonological theory. *Phonology* 11.317–323.

Marcus, Stephen Michael. 1981. Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics* 30.247–256.

Martin, Andrew. 2007. *The evolving lexicon*. Doctoral Dissertation, University of California, Los Angeles.

McCarthy, John. 2000. The prosody of phase in Rotuman. *Natural Language and Linguistic Theory* 18.147–197.

McCarthy, John. 2003. OT constraints are categorical. *Phonology* 20.75–138.

McCarthy, John, and Alan Prince. 1995. Faithfulness and reduplicative identity. *Papers in Optimality Theory*, ed. by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, volume 18 of *University of Massachusetts Occasional Papers in Linguistics*, 249–384. University of Massachusetts, Amherst: Graduate Linguistics Student Association.

McLennan, G. R. 1978. The longum and biceps of the Greek hexameter. *Mnemosyne* 31.68–70.

Mester, Armin. 1994. The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12.1–61.

Metropolis, N., and S. Ulam. 1949. The Monte Carlo method. *Journal of the American Statistical Association* 44.335–41.

Miyashita, Mizuki. 2002. *Tohono O'odham syllable weight: Descriptive, theoretical and applied aspects*. Doctoral Dissertation, University of Arizona.

Mohanan, Tara. 1989. Syllable structure in Malayalam. *Linguistic Inquiry* 20.589–625.

Mohri, Mehryar, Fernando Pereira, and Michael Riley. 2009. AT&T FSM Library. Software package, available at `www.research.att.com/~fsmtools`.

Moore, Brian. 1989. *An introduction to the psychology of hearing*. San Diego, CA: Academic Press.

Morén, Bruce T. 1999. *Distinctiveness, coercion, and sonority*. Doctoral Dissertation, University of Maryland, College Park.

Morén, Bruce T. 2000. The puzzle of Kashmiri stress: implications for weight theory. *Phonology* 17.365–396.

Mori, Yoko. 2002. Lengthening of Japanese monomoraic nouns. *Journal of Phonetics* 30.689–708.

Morton, John, Steve Marcus, and Clive Frankish. 1976. Perceptual centers (P-centers). *Psychological Review* 83.405–408.

Murugan, V. 2000. *Tolkappiyam in English*. Chennai, India: Institute of Asian Studies.

Nagarajan, H. 1995. Gemination of stops in Tamil: implications for the phonology-syntax interface. *UCL Working Papers in Linguistics* 7.495–509.

Nagy, Gregory. 2010. Language and meter. *A companion to the Ancient Greek language*, ed. by Egbert J. Bakker. West Sussex, U.K.: Wiley-Blackwell Publishers.

Narayanan, Shrikanth, Dani Byrd, and Abigail Kaun. 1999. Geometry, kinematics, and acoustics of Tamil liquid consonants. *Journal of the Acoustical Society of America* 106.1993–2007.

Niang, Mamadou. 1995. Syllable 'sonority' hierarchy and Pulaar stress: a metrical approach. *Kansas Working Papers in Linguistics* 20.53–68.

Niklas, Ulrike. 1988. *Introduction to Tamil prosody*. Bulletin de l'Ecole française d'Extrême-Orient 77.

Ohala, Manjari. 1983. *Aspects of Hindi phonology*. Delhi: Motilal Banarsidass.

Ohsiek, D. 1978. Heavy syllables and stress. *Syllables and segments*, ed. by Alan Bell and Joan B. Hooper. Amsterdam: North-Holland Publishing Co.

Oldenberg, Hermann. 1888. *Die Hymnen des Rigveda. Herausgegeben von Hermann Oldenberg. Band I. Metrische und textgeschichtliche Prolegomena zu einer kritischen Rigveda-Ausgabe*. Berlin: Hertz. Reprinted 1982 by Steiner, Wiesbaden.

Paradis, Carole. 1997. Constraint effects in Gere: From cycles to derivations. *Derivations and constraints in phonology*, ed. by I. Roca. Oxford: Oxford University Press.

Parker, Steve. 2002. *Quantifying the sonority hierarchy*. Doctoral Dissertation, University of Massachusetts, Amherst.

Pater, Joe. 2000. Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology* 17.237–274.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33.999–1035.

Pater, Joe. Forthcoming. Serial Harmonic Grammar and Berber syllabification. *Prosody matters: Essays in honor of Lisa O. Selkirk*, ed. by Toni Borowsky, Shigeto Kawahara, Takahito Shinya, and Mariko Sugahara. London: Equinox Press.

Peperkamp, Sharon. 1997. *Prosodic words*. Doctoral Dissertation, University of Amsterdam. Published as HIL Dissertations 34, The Hague, Academic Graphics.

Petrunicheva, Z. N. 1960. *Iazyk Telugu*. Moskva: Izdatel'stvo Vostochnoi Literaturi.

Pharr, Clyde. 1964. *Vergil's Aeneid books I-IV: revised edition*. Lexington, Massachusetts: D. C. Heath and Company.

Pickett, J. P. 2000. *The American Heritage Dictionary of the English language*. Houghton Mifflin.

Pipping, Hugo. 1903. *Bidrag till eddametriken*, volume 50 of *Skrifter utgifna af svenska litteratursällskapet i Finnland*. Helsingfors: Tidnings- & Tryckeri-Aktiebolagets Tryckeri.

Pipping, Hugo. 1937. Studier i metrik och prosodi. *Studier i nordisk filologi* 27.

Podesva, Robert J. 2002. Segmental constraints on geminates and their implications for typology. Paper presented at the Annual Meeting of the Linguistic Society of America.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt, and Michael Becker. 2010. Harmonic grammar with linear programming: From linear systems to linguistic typology. *Phonology* 27.1–41.

Prince, Alan. 1989. Metrical forms. *Rhythm and meter*, ed. by Paul Kiparsky and Gilbert Youmans, volume 1 of *Phonetics and Phonology*, 45–81. San Diego, CA: Academic Press.

Prince, Alan. 1999. Paninian relations. Handout, University of Marburg. Available July 2010 at `http://ling.rutgers.edu/people/faculty/prince.html`.

Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in Generative Grammar*. Blackwell. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published by Blackwell, 2004.

Probert, Philomen. 2002. On the prosody of Latin enclitics. *Oxford University Working Papers in Linguistics, Philology and Phonetics*, ed. by A. Hartmann, I. J. an Willi, volume vii, 181–206. Oxford: Oxford University, Centre for Linguistics, Philology and Phonetics.

Probert, Philomen. 2010. Phonology. *A companion to the Ancient Greek language*, ed. by Egbert J. Bakker. West Sussex, U.K.: Wiley-Blackwell Publishers.

Quené, Hugo, and Huub van den Bergh. 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* 59.413–425.

R Development Core Team. 2009. R: A language and environment for statistical computing. ISBN 3-900051-07-0, Vienna, Austria.

Rajam, V. S. 1992. *A reference grammar of classical Tamil poetry*. Philadelphia, Pennsylvania: American Philosophical Society.

Raven, D. S. 1962. *Greek metre*. London: Faber and Faber.

Ripley, B. D. 2011. MASS package for R. URL cran.r-project.org/web/packages/MASS/.

Robert, C. P., and G. Casella. 2004. *Monte Carlo statistical methods*. New York: Springer, 2nd edition.

Rubinstein, R. Y., and D. P. Kroese. 2007. *Simulation and the Monte Carlo method*. John Wiley & Sons, 2nd edition.

Russom, Geoffrey. 1998. *Beowulf and Old Germanic metre*. Cambridge, U.K.: Cambridge University Press.

Ryan, Kevin M. 2003. On the letter *āytam* in ancient Tamil. B.A. thesis in South/Southeast Asian Studies, University of California, Berkeley.

Ryan, Kevin M. 2007. Two dimensions of correspondence in half-rhyme. Paper presented at Poetics Fest II, University of California, Santa Cruz.

Ryan, Kevin M. 2010. Variable affix order: grammar and learning. *Language* 86.758–

791.

Ryan, Kevin M. Forthcoming. Contextual and non-contextual prosodic minimality. *NELS 41: Proceedings of the 41st Meeting of the North Eastern Linguistics Society*, ed. by Lena Fainleib, Nicholas LaCara, and Yangsook Park. University of Massachusetts, Amherst: GLSA Publications.

Sadeniemi, M. 1951. *Die Metrik des Kalevala-Verses*. Helsinki: Folklore Fellows Communications.

Sale, William Merritt. 1993. Homer and the Roland: The Shared Formular technique, Part II. *Oral Tradition* 8.381–412.

Sapir, Edward, and Morris Swadesh. 1960. *Yana dictionary*. Berkeley, California: University of California Publications in Linguistics 22.

Selkirk, Elisabeth O. 1982. The syllable. *The structure of phonological representations, part II*, ed. by Harry van der Hulst and Norval Smith, 337–384. Dordrecht: Foris.

Shelton, Michael. 2007. *An experimental approach to syllable weight and stress in Spanish*. Doctoral Dissertation, The Pennsylvania State University.

Sievers, Eduard. 1893. *Altgermanische Metrik*. Halle: Niemeyer.

Smolensky, Paul, and Géraldine Legendre. 2006. *The harmonic mind: From neural computation to Optimality-Theoretic grammar*. Cambridge, Massachusetts: Massachusetts Institute of Technology Press.

Steriade, Donca. 1982. *Greek prosodies and the nature of syllabification*. Doctoral Dissertation, Massachusetts Institute of Technology.

Steriade, Donca. 1990. Gestures and autosegments: Comments on Browman and Goldstein's paper. *Papers in laboratory phonology 1: Between the grammar and physics of speech*, ed. by John Kingston and Mary Beckman, 382–397. Cambridge, U.K.: Cambridge University Press.

Steriade, Donca. 1991. Moras and other slots. *Proceedings of the Formal Linguistics Society of Midamerica*, volume 1, 254–280.

Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology boundary. *Papers in laboratory phonology V: acquisition and the lexicon*, ed. by

M. B. Broe and J. B. Pierrehumbert, 313–334. Cambridge, U.K.: Cambridge University Press.

Steriade, Donca. 2001a. Directional asymmetries in place assimilation. *The role of speech perception in phonology*, ed. by Elizabeth Hume and Keith Johnson, 219–250. New York: Academic Press.

Steriade, Donca. 2001b. The phonology of perceptibility effects: the P-map and its consequences for constraint organization. MS, University of California, Los Angeles.

Steriade, Donca. 2008a. Metrical evidence for an interlude theory of weight. Handout from the City University of New York (CUNY) Conference on the Syllable, New York City, January 2008.

Steriade, Donca. 2008b. Resyllabification in the quantitative meters of Ancient Greek: Evidence for an Interval Theory of Weight. MS, Massachusetts Institute of Technology.

Steriade, Donca. 2009. Units of representation for linguistic rhythm. Slides from the Lingiustic Society of America Institute at Berkeley, August 2009.

Steriade, Donca. 2011. Rhyming evidence for intervals. Colloquium presentation at the UCLA Department of Linguistics, March 4.

Stevens, S. S. 1946. On the theory of scales of measurement. *Science* 103.677–680.

Suárez, Jorge A. 1983. *La lengua tlapaneca de Malinaltepec*. Mexico City: Universidad Nacional Autónoma de México.

Takagi, N., and V. Mann. 1994. A perceptual basis for the systematic phonological correspondences between Japanese loan words and their English source words. *Journal of Phonetics* 22.343–356.

Tarlinskaja, Marina. 1976. *English verse: Theory and history*. The Hague: Mouton.

Tarlinskaja, Marina, and L. M. Teterina. 1974. Verse-prose-metre. *Linguistics* 129.63–86.

Thinnappan, S. P. 1976. Alapetai in Tamil. *Dravidian linguistics*, ed. by S. Agesthialingom and P. S. Subrahmanyam, volume 5, 539–560. Annamalai, India: Annamalai University Press.

Topintzi, Nina. 2005. Onset weight in Arabela and Bella Coola. Paper presented at the 13th Manchester Phonology Meeting, Manchester, U.K., 26–28 May 2005.

Topintzi, Nina. 2008. On the existence of moraic onset geminates. *Natural Language and Linguistic Theory* 26.147–184.

Topintzi, Nina. 2010. *Onsets: Suprasegmental and prosodic behaviour*. Cambridge University Press.

Varian, Hal. 2005. Bootstrap tutorial. *Mathematica Journal* 9.768–775.

Velankar, Hari Damodar. 1949. *Jayadāman: A collection of ancient texts on Sanskrit prosody and a classified list of Sanskrit meters with an alphabetical index*. Bombay: Haritoṣamālā.

Venables, W. N., and B. D. Ripley. 2002. *Modern applied statistics with S*. New York: Springer, fourth edition.

Wall, Larry. 2010. Perl programming language. Version 5.12, URL `www.perl.org`.

Warren, R. M. 1970. Elimination of biases in loudness judgements of pure tones. *Journal of the Acoustical Society of America* 48.1397–1413.

Weiss, Michael. 2009. *Outline of the historical and comparative grammar of Latin*. Beech Stave Press.

West, M. L. 1970. A new approach to Greek prosody. *Glotta* 48.185–194.

West, M. L. 1982. *Greek metre*. Oxford: Clarendon Press.

West, M. L. 1987. *Introduction to Greek metre*. Oxford University Press.

Whitney, William Dwight. 1889. *Sanskrit grammar, including both the Classical language and the older dialects, of Veda and Brahmana*. Cambridge, Massachusetts: Harvard University Press, 2nd edition.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30.945–982.

Wilson, Colin, and Benjamin George. 2008. Maxent grammar tool. Software package. University of California, Los Angeles.

Wolff, Ekkehard. 1983. *A grammar of the Lamang language*. Glückstadt: Verlag J. J. Augustin.

Woodbury, Anthony. 1987. Meaningful phonological processes: A consideration of Central Alaskan Yupik Eskimo prosody. *Language* 63.685–740.

Wu, C. F. Jeff. 1983. On the convergence properties of the EM algorithm. *Annals of Statistics* 11.95–103.

Youmans, Gilbert. 1989. Milton's meter. *Rhythm and meter: Phonetics and phonology, volume I*, ed. by Paul Kiparsky and Gilbert Youmans, 341–379. San Diego, CA: Academic Press.

Zec, Draga. 1988. *Sonority constraints on prosodic structure*. Doctoral Dissertation, Stanford University.

Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12.85–129.

Zec, Draga. 2003. Prosodic weight. *The syllable in Optimality Theory*, ed. by Caroline Féry and R. van de Vijer. Cambridge, U.K.: Cambridge University Press.

Zhang, Jie. 2007. Constraint weighting and constraint domination: A formal comparison. *Phonology* 24.433–459.

Zvelebil, Kamil. 1989. *Classical Tamil prosody: an introduction*. Madras, India: Hoe & Co.