Cambridge Studies in Linguistics 125

Onsets

Suprasegmental and Prosodic Behaviour

Nina Topintzi

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521493352

This page intentionally left blank

ONSETS

The concept of the 'onset', i.e. the consonant(s) before the vowel of a syllable, is critical within phonology. While phonologists have examined the segmental behaviour of onsets, their prosodic status has instead been largely overlooked. In fact, most previous accounts have stipulated that onsets are insignificant when it comes to the 'heaviness' of syllables. In this book Nina Topintzi presents a new theory of onsets, arguing for their fundamental role in the structure of language both in the underlying and surface representation, unlike previous assumptions. To capture the weight behaviour of onsets, a novel account is proposed that relates their interaction with voicing, tone and stress. Using numerous case-studies and data from a variety of languages and phenomena (including stress, compensatory lengthening, gemination and word minimality), the book introduces a model that reflects the true behaviour of onsets, demonstrating profound implications for syllable and weight theories.

NINA TOPINTZI is a Teaching Fellow in the School of English at Aristotle University of Thessaloniki.

In this series

80. ANDREA MORO: *The raising of predicates: predicative noun phrases and the theory of clause structure*
81. ROGER LASS: *Historical linguistics and language change*
82. JOHN M. ANDERSON: *A notional theory of syntactic categories*
83. BERND HEINE: *Possession: cognitive sources, forces and grammaticalization*
84. NOMI ERTESCHIK-SHIR: *The dynamics of focus structure*
85. JOHN COLEMAN: *Phonological representations: their names, forms and powers*
86. CHRISTINA Y. BETHIN: *Slavic prosody: language change and phonological theory*
87. BARBARA DANCYGIER: *Conditionals and prediction*
88. CLAIRE LEFEBVRE: *Creole genesis and the acquisition of grammar: the case of Haitian creole*
89. HEINZ GIEGERICH: *Lexical strata in English*
90. KEREN RICE: *Morpheme order and semantic scope*
91. APRIL MCMAHON: *Lexical phonology and the history of English*
92. MATTHEW Y. CHEN: *Tone Sandhi: patterns across Chinese dialects*
93. GREGORY T. STUMP: *Inflectional morphology: a theory of paradigm structure*
94. JOAN BYBEE: *Phonology and language use*
95. LAURIE BAUER: *Morphological productivity*
96. THOMAS ERNST: *The syntax of adjuncts*
97. ELIZABETH CLOSS TRAUGOTT AND RICHARD B. DASHER: *Regularity in semantic change*
98. MAYA HICKMANN: *Children's discourse: person, space and time across languages*
99. DIANE BLAKEMORE: *Relevance and linguistic meaning: the semantics and pragmatics of discourse markers*
100. IAN ROBERTS AND ANNA ROUSSOU: *Syntactic change: a minimalist approach to grammaticalization*
101. DONKA MINKOVA: *Alliteration and sound change in early English*
102. MARK C. BAKER: *Lexical categories: verbs, nouns and adjectives*
103. CARLOTA S. SMITH: *Modes of discourse: the local structure of texts*
104. ROCHELLE LIEBER: *Morphology and lexical semantics*
105. HOLGER DIESSEL: *The acquisition of complex sentences*
106. SHARON INKELAS AND CHERYL ZOLL: *Reduplication: doubling in morphology*
107. SUSAN EDWARDS: *Fluent aphasia*
108. BARBARA DANCYGIER AND EVE SWEETSER: *Mental spaces in grammar: conditional constructions*
109. HEW BAERMAN, DUNSTAN BROWN AND GREVILLE G. CORBETT: *The syntax-morphology interface: a study of syncretism*
110. MARCUS TOMALIN: *Linguistics and the formal sciences: the origins of generative grammar*
111. SAMUEL D. EPSTEIN AND T. DANIEL SEELY: *Derivations in minimalism*
112. PAUL DE LACY: *Markedness: reduction and preservation in phonology*
113. YEHUDA N. FALK: *Subjects and their properties*
114. P. H. MATTHEWS: *Syntactic relations: a critical survey*
115. MARK C. BAKER: *The syntax of agreement and concord*
116. GILLIAN CATRIONA RAMCHAND: *Verb meaning and the lexicon: a first phase syntax*
117. PIETER MUYSKEN: *Functional categories*
118. JUAN URIAGEREKA: *Syntactic anchors: on semantic structuring*
119. D. ROBERT LADD: *Intonational phonology second edition*
120. LEONARD H. BABBY: *The syntax of argument structure*
121. B. ELAN DRESHER: *The contrastive hierarchy in phonology*
122. DAVID ADGER, DANIEL HARBOUR AND LAUREL J. WATKINS: *Mirrors and microparameters: phrase structure beyond free word order*
123. NIINA NING ZHANG: *Coordination in syntax*
124. NEIL SMITH: *Acquiring phonology*
125. NINA TOPINTZI: *Onsets: suprasegmental and prosodic behaviour*

Earlier issues not listed are also available

CAMBRIDGE STUDIES IN LINGUISTICS

General Editors: P. AUSTIN, J. BRESNAN, B. COMRIE, S. CRAIN, W. DRESSLER, C. J. EWEN, R. LASS, D. LIGHTFOOT, K. RICE, I. ROBERTS, S. ROMAINE, N. V. SMITH

Onsets: Suprasegmental and Prosodic Behaviour

ONSETS

SUPRASEGMENTAL AND PROSODIC BEHAVIOUR

NINA TOPINTZI
Aristotle University, Thessaloniki

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521493352

First published in print format 2010

ISBN-13 978-0-511-74088-6 eBook (Adobe Reader)

ISBN-13 978-0-521-49335-2 Hardback

Contents

Preface

Theme

As the title communicates, this book is about onsets. In particular, it focuses on their suprasegmental and prosodic behaviour. This on its own is quite interesting, given that most current phonological theory assigns no such role to onsets. The core of the discussion is devoted to proving this assertion wrong.

The pivotal proposal this book wishes to make is that onset weight exists. And like rimal weight, it participates in the same phenomena, that is, stress, compensatory lengthening, word minimality and gemination, among others. However, onset weight is not unrestricted. It may either be produced on the surface as a response to a weight-inducing requirement or it may be lexical, in which case its source is found in the underlying representation. The former type of onset weight is termed *coerced*, whereas the latter is dubbed *distinctive*, following Morén's (2001) similar distinction for coda weight. The first type of weight – but not the latter – is subject to certain limitations and thus can be assigned to a subset of segments only.

These restrictions represent a leading idea in the book. In particular, it is claimed that the optimal weightful onsets are those that lack the feature [+voice]. As a result, the prototypical moraic onsets are [−voice] ones, i.e. the voiceless obstruents. Justification for this account comes from the relationship between voicing, tone and stress. Examination of the tonogenesis facts reveals that the pitch perturbation due to (the lack of) voicing is commonly phonologized as tone. Extending this idea, I propose that in some languages such pitch perturbation is phonologized as stress, in terms of moras. The languages Karo, Pirahã and Arabela exhibit this pattern.

Of paramount importance is a related claim made with regard to sonorants. These are argued to be marked on a language-specific basis as [+voi] or lack any [voi] feature whatsoever. This statement is supported by the behaviour of sonorants in the examination of tone and stress. It is postulated that they may pattern with the voiced obstruents when they bear [+voi]; but they may also

pattern alongside the voiceless obstruents, if they lack [voi] specification. As explained, the system built is such that it can treat [−voi] segments and those lacking [voi] uniformly to the exclusion of [+voi] ones.

Still, as far as onset-sensitive stress is concerned, a crucial distinction is made. Languages may be sensitive to the presence of an onset (PO effect) or to its quality (QO effect), but also to both (PO&QO effect). Importantly, PO effects are down to alignment considerations, whereas QO effects depend on weight considerations. For instance, in Aranda, an onsetful syllable attracts stress more than an onsetless one (PO); in Karo, syllables with onsets of a certain quality attract stress more than others (QO); and finally, in Pirahã, the two effects are combined (PO&QO).

The subsequent discussion centres on the interaction of onsets with other phenomena, namely compensatory lengthening, word minimality and gemination. The reasoning behind this is the following: onset-sensitive stress on its own is not sufficient to uphold the onset-weight hypothesis. This is because there are other ways to account for stress, e.g. the concept of prominence, which bypass reference to weight. Thus, providing evidence on the effects of onsets from uncontroversially weight-based phenomena surely offers solid grounding to the onset-weight theory.

To this end, a large number of case-studies is examined and formalized using the framework of Optimality Theory (OT) (Prince and Smolensky 1993/2004; McCarthy and Prince 1995). Many of the analyses are very detailed, thus also supplying the technical implementation of the theoretical ideas proposed here.

Finally, the theoretical proposals that are presented often extend beyond onsets. For instance, the chapters on compensatory lengthening and geminates provide full-fledged analyses (in terms of theoretical and empirical scope, as well as technical detail) that promise to account for a wide range of facts and cases.

Audience and use

This book revolves around the topic of onsets, but in doing so, it addresses several core phenomena in suprasegmental and prosodic phonology. It is thus of interest to anyone who works on syllables, geminates, weight theory, compensatory lengthening, word minimality, tone and reduplication. It will also prove useful to segmental phonologists, especially those interested in [voice] and its interaction with the prosodic phenomena of tone and stress.

Typologists interested in phonology can certainly use this book as a resource on some typologically rare languages and find information on languages hardly discussed elsewhere.

This study should be accessible to anyone with some background in phonology. While it is true that its technical aspects will be fully appreciated by OT-theorists (or anyone with some good knowledge of OT), the core insights, such as the weightfulness of onsets and the relationship between stress and voicing, are theory neutral. The book can be used in various ways: by the phonologist in general, as a new model of syllable and weight to apply in her/his work; by the OT-theorist who wishes to espouse (some of) the ideas proposed here for their technical merits; by the typologist, as a resource for typological studies; and by the student, as a useful review of past and recent theories on syllable weight, compensatory lengthening and gemination.

A final word

This book is an updated and revised version of my 2006 thesis titled 'Moraic Onsets', at University College London. Much of the material in here is significantly altered, although the core idea remains the same: 'onsets can be moraic'! Most analyses have been changed (e.g. Arabela, Bella Coola, Samothraki Greek, etc.), while the discussion on [voice] as well as sonorants has been updated to reflect findings of more recent work, such as Tang (2008). Some material has been discarded in the interest of clarity, while certain other sections have been added (see the discussion on medial-onset geminates). The interested reader can consult Topintzi (2006b) for comparison. Finally, earlier versions of some of the material presented here have appeared in previously published work of mine. This includes: Topintzi (2006a) and Topintzi (2008a, b).

Acknowledgements

In December 2001, as an MA student at UCL, I listened to a colloquium talk by Dan Everett on the phonology of Pirahã. I thought Dan was an engaging speaker and the data he reported on fascinating. A year later, as a first-year PhD student, I thought about this talk again; Pirahã has been notorious for its onset-sensitive stress. But, was it the only such language? Could there be more cases like it, and in fact, could it be that the role of onsets in prosody had so far been underestimated? These questions persist within this book – and previously in my thesis – in the hope that by the end of it a satisfactory answer will have been given.

During the years that I've been looking into this topic (2002–9), a number of people generously contributed to it with their suggestions, comments and help in gaining access to relevant material. First and foremost, my thanks go to Moira Yip, my thesis supervisor and academic role model, for teaching me good phonology (any 'bad' phonology here is purely my fault), helping me shape my own ideas, sharing her expertise and vast knowledge and for the most exquisite 'Tosca' I have and will possibly ever enjoy. Dan Everett has been a constant resource regarding Pirahã and provided a lot of encouragement when I first took up this project. Since 2005, Marc van Oostendorp has been my 'life-vest' whenever I seemed to get stuck on a piece of data, always offering insightful ways to tackle the problem in question. I thank both a lot! Many thanks also go to Matt Gordon, Iggy Roca and a third anonymous reviewer for this book. I thank them for their careful reading and their invaluable comments. I am especially grateful to Matt, who offered the most helpful feedback possible, disregarding his own views on the issue – expressed in Gordon (2005) – and generously offering remarks to make my own work genuinely better.

Besides the above, I also gratefully acknowledge: Mary Baltazani, Jean-Marc Beltzung, Juliette Blevins, Lev Blumenfeld, Emily Curtis, Paul de Lacy, Mark Donohue, Paloma Garcia-Bellido, Spike Gildea, Rob Goedemans, Carolina González, Beverley Goodman, Carlos Gussenhoven, John Hajek,

John Harris, Bruce Hayes, Ben Hermans, Larry Hyman, Wouter Jansen, Brian Jose, Dasha Kavitskaya, Michael Kenstowicz, Debbie Koop (SIL Peru), Yen-Hwei Lin, Angeliki Malikouti-Drachman, Bruce Morén, Elliott Moreton, David Nash, Mary Pearce, Anthi Revithiadou, Curt Rice, Keren Rice, Roland Rich, Joe Salmons, Toshi Shiraishi, Jen Smith, Paul Smolensky, Donca Steriade, Mary Ruth Wise, Yi Xu and Hideki Zamma. Thanks to Eleftheria Giakoumaki and Angelos Afroudakis at the Κέντρον Ερεύνης των Νεοελληνικών Διαλέκτων και Ιδιωμάτων (Academy of Athens Research Centre for the Greek Dialects and Idioms) – Ι.Λ.Ν.Ε for granting me access to older Samothraki Greek sources. Also to Marianna Ronga, Maria Tsolaki and Allia Foteinou, my informants on Samothraki Greek.

Warm thanks also go to my fellow students at UCL (2002–6) and my colleagues at the Universities of Patras and Thessaloniki (AUTH), where I have been employed for the most part till now (2007–9). Many thanks (and congratulations!!) to colleagues and friends for standing by and for occasionally putting up with me during the last couple of years: Xara and Tasoula Balaska, Mary Baltazani, Yannis and Labros Giogios, Ioanna Kappa, Giouli Kommata, Marika Lekakou, Despoina Papadopoulou, Vina Tsakali and Eirini Tsiknidou.

For wonderful co-operation and communication, I am grateful to Helen Barton, my editor at CUP.

I kept my warmest thanks for last. This goes without saying to my family: my parents Betty and Takis and my sister Ermina (and to our newest, furrier, family members Astreha and Akrali). Their contribution is one that this work could not have existed without: unconditional love and unfailing support.

1 *Onsets and weight: the theory*

> *. . . no language has a rule stressing the penultimate syllable unless it begins with a voiced consonant, in which case one stresses the antepenultimate syllable* Hyman 1985: 96

> *[Karo] stress can be predicted by the onset of the last syllable: if it is a voiced stop consonant, then the stress shifts one syllable to the left* Gabas 1999: 39

1.1 Aims

This book explores the role of onsets in syllable-weight theory and consequently on prosody. In particular, I will be arguing that onsets, like nuclei and codas, can bear weight. Although this idea is one that has to be seriously considered by every phonologist interested in stress, weight and prosodic structure, it has nonetheless been overlooked. In fact, virtually all work on syllable structure and weight, the most prominent being Hyman (1985), Hayes (1989), Goedemans (1998), Morén (2001) and Gordon (2006, the published version of his 1999 thesis), maintains that onsets can never be weightful. However, in most cases, this claim is made purely by stipulation and for convenience, as e.g. Morén (2001: 8) also acknowledges: 'Onsets are typically non-moraic. Although this is not the only logical possibility, it is *convenient* and I assume it here' (emphasis added mine). The present book aims to fill this gap in the literature and challenge the standard assumption that sees onsets as weightless, offering fresh insights around this topic.

This comes in support of the recent marginal literature – basically Hajek and Goedemans (2003) and Gordon (2005) – that hesitantly admits some type of weightful onsets. However, unlike those accounts that focus on specific aspects of the issue, e.g. stress (Gordon 2005) or geminates (Hajek and Goedemans 2003), this book offers the first comprehensive study in terms of argumentation, length and concreteness on the issue of onset weight.

Using findings of the past literature as a starting point, it presents a novel theory whose ambition is to encompass phenomena from a host of languages (many of which had never been discussed in this light before) in a unifying, explanatory and restrictive way. It argues against the prosodic inertness of onsets and proposes a modification of the syllable-weight model where onsets, like codas, can sometimes be moraic and sometimes not. To this end, the book may also be used as a resource for other researchers, as it contains a collection of languages and data which have been argued (rightly or not) to exhibit onset sensitivity.

This first chapter consists of two parts. The first (§1.2) is quite introductory. It aims at establishing the need for a book of this kind and sets the scene for the topic under consideration. The second, longer, part (§1.3) is more theoretical and technical in nature. It supplies the theory that will be implemented in the forthcoming chapters where several case-studies will be discussed and analysed.

1.2 Why a book on moraic onsets?

This part starts by briefly looking into syllable weight (§1.2.1), as well as some of the models that have been proposed to capture it. It will soon become evident that moraic theory (Hyman 1985; Hayes 1989) stands out as the most successful of all (§.1.2.2). In its standard conception, however, moraic theory proves empirically insufficient, since it explicitly excludes a range of cases and data that are actually attested (§1.2.3–4). Rather than replacing it, though, this book argues that with some modification – namely by allowing onsets to be weightful – moraic theory can incorporate these cases too, and thus emerge as a complete, accurate and yet restrictive theory of weight. Section 1.3 explains how this is possible.

1.2.1 Syllable weight

Syllable weight refers to the idea that syllables of different structure behave in different ways in prosodic phenomena and processes such as stress, reduplication, tone, compensatory lengthening, word minimality and others. In the languages that make a distinction based on weight, syllables are either heavy or light. They are heavy if they contain a long vowel (VV) and light if they are simply made up of a short vowel (V). Syllables with a short vowel followed by a coda (VC) are heavy or light depending on the language. In many languages stress is attracted to heavy syllables. Hopi and Lenakel are languages of this sort. However, while in Hopi VC counts as heavy, in Lenakel it counts as light.

(1) *Hopi: VV/VC=heavy; V=light*

a.	qǿq.tø.som.pi	‘headbands’
	só:.ja	‘planting stick’
b.	qø.tǿ.som.pi	‘headband’
	ko.jó.ŋo	‘turkey’

In Hopi (Jeanne 1982, cited in Gussenhoven and Jacobs 2005: 145), the first syllable is stressed if it is heavy (C)VV or (C)VC (1a), but if it is light (C)V, then the second syllable gets stress (1b).[1] In Lenakel (Lynch 1974, 1978 cited in Hayes 1995), on the other hand, (C)VCs are considered light and primary stress appears on the penult (2a). Simplifying a bit, secondary stress is (usually) assigned to the first syllable and to alternate syllables after that (2b). However, this pattern may be disrupted; (C)VCs do not get secondary stress (cf. unstressed *mol* in (2d)) unless they happen to be located in a position that would receive rhythmic stress anyway; in contrast, heavy (C)VVs get stress no matter what their position (cf. *kì:* in 2c).

(2) *Lenakel: VV=heavy; VC/V=light*

a.	éheŋ	‘to blow the nose’	(Lynch 1978: 16)
	rɨmáwŋɨn	‘he ate’	(Lynch 1978: 19)
b.	lètup^{w}àlukáluk	‘in the lungs’	(Lynch 1974: 183)
c.	nɨkì:nílar	‘their (pl.) hearts’	(Lynch 1974: 198)
d.	rɨmolkéykey	‘he liked it’	(Lynch 1978: 19)

Several other phenomena make reference to syllable weight. A by-no-means-exhaustive list includes:

(3) *Phenomena involving syllable weight*

I. Compensatory lengthening: the lengthening that occurs after deletion of a segment
e.g. Turkish (Roca and Johnson 1999) *tahsil* → *ta:sil* ‘education’

II. Word minimality: the minimum word size some languages impose [commonly (C)VC or (C)VV]
e.g. Dalabon (Capell 1962; Garrett 1999) words that are CVC *bad* ‘stone’ or CVV *bi:* ‘man’ are allowed, CV words are banned

III. Poetic metre: the organization of syllables into feet in songs or poetry
e.g. Greek and Latin dactylic hexameter in epic poetry. The verse consists of six metra, each of which is made up of one heavy and two short syllables (–⏑⏑); however, two short syllables can be replaced by one heavy in which case we have a spondee (– –). Boundaries of metra are marked by parentheses:

[1] In this chapter, *unless stated otherwise*, the acute accent marks primary stress, the grave accent means secondary stress and underlining denotes the reduplicated portion. I will interchangeably use VV or V: to refer to long bimoraic vowels.

(cārmĭnă) (quāe vūl)(tīs cōg)(nōscĭtĕ); (cārmĭnă) (vōbīs) (Vergil, *Eclogues* VI. 25)

IV. Reduplication: the repetition of part of a word (commonly a heavy syllable)
e.g. Mokilese progressive (Harrison 1976; McCarthy and Prince 1986) *poki~pokpoki* 'beat', *kookɔ~kookookɔ* 'grind coconut', but *pa~paapa* 'weave'

V. (Prosodic) Truncation: shortening of forms, as in the production of nicknames
e.g. among other patterns, acceptable Japanese nicknames are a single heavy syllable (Mester 1990; Benua 1995) *Midori~Mii-čan* or *JuNko~JuN-čan* (*-čan* is the diminutive suffix)

VI. Gemination: the consonant doubling that often occurs after short stressed vowels, so that the syllable is rendered heavy
e.g. Kukatj (Breen 1992) or in Swedish dialects (Kiparsky 2008b) such as *viss.na* 'to wilt', *takk.sa* 'rate', *hall.va* 'half'

VII. Tone: the use of pitch to mark different morphemes
e.g. in Hausa (Gordon 2006) contour tones are only tolerated in heavy, but not in light, CV syllables, i.e. *lâ:lá:* 'indolence', *mântá:* 'forget', *râs:á:* 'branches'

The distinction between heavy and light syllables was recognized as early as Jakobson (1931) and Trubetzkoy (1939) and has since been formalized in three major ways: a) CV theory (McCarthy 1979; Clements and Keyser 1983), b) the X-slot model (Levin 1985), and c) the moraic model (Hyman 1985; Hayes 1989). All three theories assign abstract weight units to segments in the syllable. The difference lies in what kinds of units these are and exactly what syllable constituents are identified, which of course has repercussions on the predictions made. For example, the syllable *tan* would be represented in the first two models in the following way.

(4)

a. *CV theory*

b. *X-slot model*

R=rime
O=onset
N=nucleus
C=coda

In CV theory (4a), segments are marked as C-ones and V-ones, whereas in the X-slot model (4b), the more generic tag X is used to refer to both consonants and vowels. The latter notation has a welcome result; there is evidence that the C and V labels can sometimes be far too specific. For instance,

in Ancient Greek, the form *esmi* 'I am' underwent *s*-deletion and subsequent compensatory lengthening. In some dialects, the resulting form was *emmi* with C-lengthening, while in others it was *eemi* with V-lengthening. CV theory can account for *emmi*, because the vacated C-position of *s* is filled by a consonant too, but it fails to do so in *eemi* where the position is held by a V. This is not a problem shared by the X-slot theory, since X slots, by being general enough, circumvent this problem of labelling.

X slots too, however, prove inadequate. Consider the example of Japanese from (3V) above. Given hypocoristics such as *Midori~Mii-čan* or *JuNko~JuN-čan*, one can simply state that hypocoristic formation involves heavy [XXX] templates, i.e. [CVV] or [CVC] syllables. This idea cannot be maintained once other possible nicknames are considered, as shown in (5).

(5) *Japanese Hypocoristics* (Benua 1995)

Midori	Mido-čan, Mii-čan
Hanako	Hana-čan, Haa-čan, Hač-čan

In these examples, the nicknames – excluding the diminutive suffix *čan* – can either be monosyllabic [CVV] or [CVC] or bisyllabic [CVCV]. Obviously, this pattern cannot be captured by a template [XXX]. Data like these have led to a further improvement of the syllable-weight theory by utilizing the concept of moras, as proposed in Hyman (1985) and especially Hayes (1989).

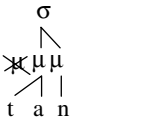
1.2.2 Advantages of moraic theory

Hyman (1985) proposes a model which consists of weight units (WUs) whose function is virtually identical to moras, which is why I will simplify and use moras for Hyman's representations too (6a). For our purposes, the most important property of this model is that underlyingly all segments start off with at least one WU (6a.i). Crucially on the surface, onsets lose their WU (indicated by the crossed-out mora in (6a.ii)) due to the universal application of the Onset-Creation Rule (OCR). This rule applies whenever a [+cons] segment is followed by a [−cons] segment and its effect is to delete the WU of the [+cons] segment. Subsequently, the [+cons] feature matrix associates to the WU of the [−cons] segment on its right. In other words, the nucleic WU/mora dominates both the onset and the nucleus of the syllable (6a.ii).

(6) a. *Hyman (1985)*

i. underlying form

ii. surface form



b. *Hayes (1989)*

surface form

This differs from Hayes (1989), who assumes that the nucleic mora is not shared between the onset and the nucleus, but only associates to the nucleus. The onset instead links directly to the syllable node as depicted in (6b). Note that although I have represented the coda consonant in (6b) as moraic, a singleton coda consonant may be non-moraic on the surface (compare Hopi (C)V$^{\mu}$C$^{\mu}$ with Lenakel (C)V$^{\mu}$C previously). If it is moraic, this is the result of the application of the Weight-by-Position rule which assigns moras on codas. Thus, monomoraic (C)V syllables are light, while bimoraic (C)VVs are heavy; (C)VC can be light or heavy on a language-specific basis. Moras are also grouped into feet (McCarthy and Prince 1986; Hayes 1995), which are part of higher prosodic structure that includes prosodic words (Selkirk 1980, 1984a; Nespor and Vogel 1986; Itô and Mester 1992). Reference to feet and moras allows us to account for numerous data, many of which cannot be adequately accounted for in other timing models.

A concrete example of this sort emerges in the consideration of the seemingly problematic data from Japanese hypocoristic formation in (5) above. To account for the attested patterns, the X-slot model needs to impose both [XXX] and [XXXX] templates; worse still, the CV model needs to utilize [CVV] or [CVC] or [CVCV] templates depending on the nickname considered each time. Evidently, none of these approaches is insightful. The same is not true for the moraic framework, which can propose a uniform and straightforward account, namely that Japanese nicknames have the size of a single bimoraic foot and consequently can emerge as either heavy CVV/CVC monosyllables or light CV disyllables.

A similarly neat explanation is available for the Ancient Greek compensatory lengthening (CL) data above. Given that CL is just about the preservation of the mora after the deletion of the segment that hosted it through lengthening of a neighbouring segment (Hayes 1989, but see Ch. 3), either V- or C-lengthening will do. Consequently, the preservation of the mora of the deleted *s* in *esmi* can yield either *eemi* or *emmi* leading to the attested dialectal variation.

In sum, moraic theory has proven more successful compared to its predecessors, and because of this, it will be the timing model assumed in this work.

1.2.3 The traditional stance of moraic theory towards onsets, and its problems

Moraic theory à la Hayes (1989) claims that one of its strong points is that it does not allow moraic onsets. Hayes argues that onset consonant deletion

never causes CL, which is only natural if onsets never bear weight. Notably, this effect cannot be as easily derived in the previous frameworks, whereby a timing slot is assigned to the onset, implying, at least in principle, that this constituent too may be active in prosodic processes. One way that has been utilized to avoid this superficially implausible result was to introduce the rimal node (cf. (4b)) as the only one that could bear weight. Moraic theory is advantageous in that respect too. Given that it only assigns moras to nuclei and codas, it can dispense with the rimal node.

Despite its virtues, moraic theory bases its argument against the existence of weightful onsets on the fallacious argument of rarity; as Hayes (1995: 51) puts it: 'Onset segments are prosodically inert… While this claim is not fully valid at the observational level, it is so well supported across languages that it serves as the central observation for formal theories of syllable weight'. Similar statements are made by other researchers too, who admit that claiming that onsets never play any role in weight is not entirely accurate. For example, Gordon (2006: 3) observes that 'in Latin, as in *virtually* all languages, the onset is ignored for purposes of calculating weight' (emphasis added mine). The bottom line seems to be that because the overwhelming majority of languages ignore onsets for prosodic processes or – to put it another way – because the prosody of languages so rarely pays attention to onsets, syllable-weight theories have so far ignored onsets by *stipulating* that they are prosodically inert.

The present book instead challenges the 'convenience' of the traditional assumptions and takes the position that although it is true that there is a very strong tendency for onsets not to matter for weight purposes, this is by no means universal. A more complete understanding of syllable structure and weight thus forces us to include onsets in the syllable-weight equation. This then suggests a more literal interpretation of the term 'syllable weight', which up to now has basically corresponded to rimal weight. In fact, admitting the participation of onsets in prosodic phenomena seems to be the null hypothesis, rather than excluding them as the traditional theory advocates.

I will consequently attempt to show that certain stress and syllable-weight facts cannot be re-analysed in any way other than by admitting weightful onsets, thus moraic onsets *do* exist in some languages and are represented in the way shown in (7b). Their introduction does not undermine moraic theory, but aspires to improve the range of facts that the theory can account for. Importantly, I am not suggesting that the presence of moraic onsets is unrestricted; rather, it is regulated by certain patterns pertaining to voicing (§1.3.3.1–3, §1.3.3.5) or

underlying moraic specification (§1.3.4). In this view, moraic theory remains in an advantageous position, because even after the introduction of moraic onsets, it can still distinguish between languages that have them (7b) and ones that do not (7a) by simply assigning a mora on the onset of the former but not the latter.

(7) a. *Non-moraic onsets* b. *Moraic onsets*

While the presence or lack of moraicity in moraic theory is built in (cf. the case of codas), other timing models do not have this option, stating that onsets should either be consistently weightful or weightless across languages. Neither situation reflects reality, however, as we will see in due course.

To conclude, in principle, there is nothing wrong with having a moraic onset. Consequently, within the current proposal, onsets come in two flavours: non-moraic (7a) – as in most languages – and moraic (7b). The latter's distribution is systematic and restricted, as discussed in §1.3. An onset can still be seen as the tautosyllabic prenucleic consonant.

1.2.4 Onset weight: a brief overview of the empirical data

To be able to follow the argumentation and justification of the theory that will be presented in the following sections, it is at this point important to consider briefly some of the data that provide the basis upon which it will be built. The subsequent chapters of course develop detailed case-studies of these and many more data where the full range of onset-weight effects are analysed extensively.

1.2.4.1 Onsets and stress

Our attention will first be drawn to stress. The stress algorithm of a handful of languages is sensitive to the presence and/or to the quality of an onset. These two factors are independent of one another, as I will be arguing, so it is possible that they interact or act separately (8). In languages like Karo (❶) only the onset's quality matters (QO systems); in others, e.g. Aranda (❷), only its presence does (PO systems); while in Pirahã (❸) both the presence and the quality of the onset are instrumental to stress assignment. More commonly, of course, neither of the two factors exerts any influence on stress assignment (❹). The data in (9) exemplify.

(8) *Presence and quality of onset interaction in stress*

QO	PO	Languages	Pattern Identifier
✓	×	Karo	❶
×	✓	Aranda, Banawá, Dutch	❷
✓	✓	Pirahã	❸
×	×	Greek, Russian, etc.	❹

(9) *Presence and/or quality effect of onsets on stress*

❶ Karo (stressed syllables in bold; Gabas 1998: 22, 1999: 39–41)

a. **ci**gi ‘spot’
pibɛʔ ‘foot’

b. pa**kː**ɔ ‘fish (sp.)’
na**hek**ˀ ‘fontanel’

c. ma**ŋᵍɔt**ˀ ‘again’
k**ɨɾɨwɛp**ˀ ‘butterfly’

d. pɛ.**ɔᵈn** ‘skin’
e.**i** ‘irara’[2]

❷ Aranda (accents indicate primary (acute) and secondary (grave) stress; Strehlow 1944; diacritics ignored)

a. tárama ‘to laugh’
kútuŋùla ‘ceremonial assistant’

b. ankáta ‘Jew lizard’
ulámbulàmba ‘water-fowl’

❸ Pirahã (stressed syllables in bold, acute accent = H tone, no accent = L tone; Everett and Everett 1984; Everett 1988)

a. ko.ʔo.**pa** ‘stomach’
b. poo.**gáí**.hi.aí ‘banana’
c. **ʔí**.bo.gi ‘milk’
d. bi**ísai** ~ mi**ísai** ‘red’

❹ Greek

a. pérazma ‘way-through’
perúka ‘wig’
ekpébo ‘transmit’
ekpobí ‘show’

b. étimos ‘ready-MASC-SG’
eláfi ‘deer’
eðáfi ‘land-PL’
éðrano ‘bench, desk’

In Karo, default stress is word final, unless some requirement, i.e. tone, nasalization or onset voicing, causes shift from that position. In particular, final

[2] Gabas (1998) is written in Portuguese. I have not translated the glosses into English.

voiced obstruent onsets repel stress (9❶a), whereas voiceless obstruent (9❶b) and sonorant (9❶c) onsets do not. Onsetless syllables are allowed (Gabas 1999: 24), and they can carry stress too if they make the best available stress bearers (9❶d). This suggests that onsetless syllables in Karo are not treated in any special way.

In Aranda, on the other hand, stress on onsetless syllables is avoided, so that actually the first onsetful syllable receives stress irrespective of its type (compare onsetful (9❷a) with onsetless (9❷b)). In Pirahã (Everett and Everett 1984; Everett 1988), the rightmost heaviest syllable of the final three in a word receives stress according to the following hierarchy: PVV > BVV > VV > PV > BV (P = voiceless onset, B = voiced onset, > = is heavier than). Thus, stress is final if all syllables are of the same type (9❸a), but may appear elsewhere when syllables are different. In particular, onsetful syllables attract stress more than onsetless ones (9❸b, gáí > aí), but also onsets of a certain type, i.e. voiceless obstruents, attract stress more than the voiced obstruents and sonorants (9❸c, ʔí > bo, gi; 9❸d, sai > bií, mií).

Finally, in languages such as Greek, the presence or quality of onsets plays absolutely no role in the stress algorithm. Syllables with onsets of any type may receive stress (9❹a) and onsetless syllables may carry stress too (9❹b).

1.2.4.2 Onsets and geminates

Another situation where onsets are prosodically active is in languages such as Trukese and Pattani Malay that exhibit initial geminates (represented as C_iC_i or as C: in the sources). Evidence for the contribution of onsets to weight comes from word-minimality effects (Trukese) and stress (Pattani Malay). More specifically, minimal words in Trukese are either CVV or C_iC_iV. The latter include a geminate. A straightforward account of this pattern is that the minimal word is bimoraic, thus implying that geminates must contribute a mora; given that the geminate can be plausibly syllabified in an onset position, it is reasonable to propose that Trukese allows moraic onsets.

(10) *Trukese initial geminates and minimal words* (Dyen 1965; Goodenough and Sugita 1980; Hart 1991; Davis 1999b)

a. CVV words	maa	‘behaviour’
	oo	‘omen’
b. C_iC_iV words	tto	‘clam sp.’
	kka	‘taro sp.’

Similar results obtain in Pattani Malay, which also has initial geminates as shown in (11b). Comparison of the minimal pairs in (11a) and (11b) reveals

that geminates have to be weight-contributing since they attract stress (11b). Again, an analysis that represents initial geminates as moraic onsets seems in order.

(11) *Pattani Malay initial geminates and stress* (Yupho 1989)

a.	buwɔ́h	'fruit'
	ɟalé	'street/path'
b.	bːúwɔh	'to bear fruit'
	ɟːálɛ	'to walk'

1.3 The theory of onset weight

1.3.1 Types of moraic onsets

Having provided empirical evidence on the existence of onset weight, the question that now arises is: what kind of moraic onsets are there? I would like to propose that there are two types of moraic onsets: a) distinctive and b) coerced, inspired by Morén's (2001) discussion on weight. Distinctive weight refers to phonemic weight distinctions; for instance, the difference between Hungarian *vicɛ* 'janitor' and *vic.cɛ* 'his joke' is one that will be represented in the input as /vicɛ/ vs. /vicᵘɛ/. Coerced weight refers to weight acquired in the output as a result of a requirement such as Word Minimality, Stress-to-Weight, or Weight-by-Position, etc. and where an input /C/ turns in the output into [Cᵘ]. I argue that such a distinction is not only applicable to codas, as Morén claims,[3] but to onsets too. In other words, this grouping distinguishes between: a) singleton non-moraic onsets vs. singleton moraic onsets (coerced weight) and b) singleton non-moraic onsets vs. geminate moraic ones (distinctive weight).

At this point, it is important to clarify certain terms.[4] As stated, distinctive weight refers to underlying weight, whereas coerced weight refers to derived weight on the surface. Moraic onsets like the ones appearing in Karo or Pirahã are clearly coerced (9). These appear moraic, but without any evidence for underlying moraicity and without any surface contrast with their non-moraic counterparts (i.e. we don't get both [p] and [pᵘ], but only the latter). Similarly, other consonants clearly bear distinctive weight; these must therefore be underlyingly moraic, i.e. geminates, since on the surface we find contrasts between

[3] As a matter of fact, Morén's distinctive–coerced contrast applies to both vowels and consonants, but he explicitly stipulates the absence of onset weight. If this stipulation is wrong, as I claim it is, then he too would be bound to expect a similar contrast in onsets.

[4] I am indebted to Bruce Morén (personal communication (p.c.)) for relevant discussion.

them and their singleton counterparts, e.g. Trukese [təə] 'islet' vs. [tto] 'clam' (Davis 1999b). These two cases so far are the prototypical coerced and distinctive moraic onsets, respectively.

This grouping is not always as clear-cut, though. Another type of moraic onset seems to exist that I will call here 'geminate*d*'.[5] These are consonants whose weight appears intermediate between distinctive and coerced, as they present properties of both. Consider, for instance, heavy consonants whose weight emerges as the product of reduplication or prefixation, e.g. Marshallese reduplication /korap/ > [yo**kk**oraprap] (§6.2.4.2). These are similar to geminates (distinctive weight) in that they present a contrast between non-moraic and moraic consonants in the same position, e.g. [**k**o.rap] vs. [yo.**kk**o.rap.rap]; but they also resemble coerced moraic onsets, since this contrast does not seem to be underlying.[6]

In the schema below (12), straight, solid links between distinctive and true geminates, and coerced and non-geminated ones denote the prototypical cases of each category, as explained above. For the reasons mentioned before, 'geminated' consonants are somewhere in between, but in the absence of negative evidence, have a closer link to distinctive weight (indicated by the thicker solid line), than to coerced weight (indicated by the dotted line).

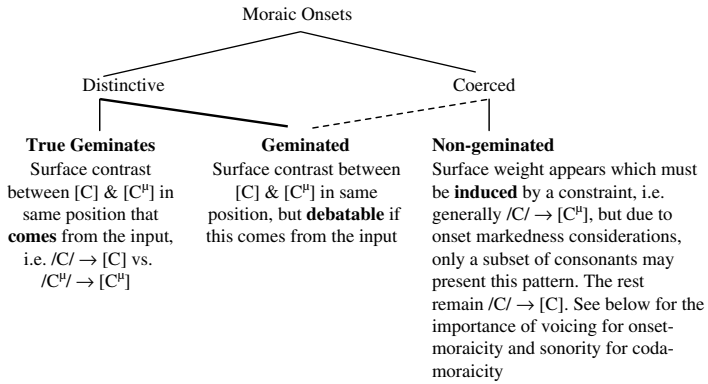
Notably, the difference between 'geminates' and 'geminated' consonants seems to be irrelevant to the forthcoming discussion (see Ch. 5 and Ch. 6), so I will no longer employ it. Instead, I will uniformly term both ***geminates***. There are two reasons for this: first, as noted above and in Chapter 5, the distinction is often difficult to determine or even undetectable; second, I will show that moraic onsets that are unambiguously distinctive or coerced largely participate in the same weight phenomena thus rendering such simplification justifiable.[7]

[5] Note that the related term 'geminate' is somewhat confusing, as it is used in the literature to refer to several things. For instance, some researchers call sequences of identical consonants 'geminates' although they lack weight (Selkirk 1990; Tranel 1991), as in e.g. Malayalam. I will claim that these are **not** geminates, but simply doubled consonants with two root nodes (cf. §5.4.2.2 for fuller discussion).

[6] The latter conclusion is, however, not uncontested (also see fn. 11 in Chapter 5). If, for instance, prefixation or reduplication involve the presence of an *underlyingly* unlinked mora that as a result of the constraint ranking ends up creating a moraic onset, we could still possibly claim that this moraic onset has distinctive weight, i.e. is a geminate, because it has a mora whose source is in the input. If this proves correct, then such consonants would need to be considered distinctive, patterning exactly like the geminates.

[7] I say 'largely' because unambiguous coerced onsets cannot be assumed in the cases of geminates, since, by definition, the latter are distinctive.

(12)



While the difference between 'geminate' and 'geminated' consonants will be disregarded, the difference in the nature of weight in the examination of prototypical cases of weight has repercussions on what kinds of moraic consonants can be found (Morén 2001). In distinctive weight there are no expectations as to what type of segments will show contrasts between singletons versus geminates. Since weight is lexically specified in the input, this will be unpredictable and consequently perhaps arbitrary. High-ranking moraic faithfulness – which I assume throughout – ensures that weight contrast of this type is preserved on the surface. But the same principle does not hold in coerced weight, which is moulded on the surface, i.e. where markedness constraints are applicable. For *nucleus* and *coda* weight, sonority considerations become relevant (Zec 1988, 1995; Morén 2001). Focusing on codas in particular, hierarchies referring to moraicity show that more sonorous segments are more likely to be moraic as well as to attract stress, leading to the conclusion that the more sonorous a coda segment is, the more likely it is for it to be moraic.

In this light, a case like Kwakwala (Zec 1988, 1995), where the moraicity of codas in CVC syllables is variable depending on whether the codas are sonorous, is an instance of coerced weight. In particular, non-glottalized sonorant codas are moraic, whereas glottalized sonorants and obstruents are not. The prediction, then, is that in languages with coerced coda weight, sonority is crucial and thus it should not be possible for non-sonorous codas to be moraic with more sonorous ones being non-moraic.

In languages with distinctive weight, however, no similar restrictions apply. Lexical specification occurs and therefore sonority reversals with respect to