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.	lètupʷàlukáluk	'in the lungs'	(Lynch 1974: 183)
c.	nɨkì:nílar	'their (pl.) hearts'	(Lynch 1974: 198)
d.	rɨ̀molkéykey	'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) *CV theory* *X-slot model*

a.


b.


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

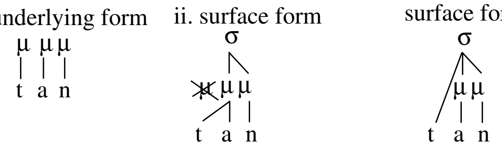
In these examples, the nicknames – excluding the diminutive suffix *čan* – can either be monosyllabic [CVV] or [CVC] or bisyllabic [CVCV]. Obviously, this pattern cannot be captured by a template [XXX]. Data like these have led to a further improvement of the syllable-weight theory by utilizing the concept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

i. underlying form ii. surface form 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^{μ}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

Cambridge University Press
978-0-521-49335-2 - Onsets: Suprasegmental and Prosodic Behaviour
Nina Topintzi
Excerpt
More information

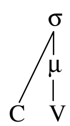
never causes CL, which is only natural if onsets never bear weight. Notably, this effect cannot be as easily derived in the previous frameworks, whereby a timing slot is assigned to the onset, implying, at least in principle, that this constituent too may be active in prosodic processes. One way that 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.

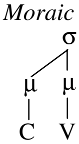
The present book instead challenges the ‘convenience’ of the traditional assumptions and takes the position that although it is true that there is a very strong tendency for onsets not to matter for weight purposes, this is by no means universal. A more complete understanding of syllable structure and 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**k:ɔ** 'fish (sp.)'
na**hek**˺ 'fontanel'
c. ma**ŋᵍɔt**˺ 'again'
kɨ**rɨwɛp**˺ '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. bií**sai** ~ mií**sai** '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