As alive as ever: the geminate debate under Containment

Nina Topintzi & Eva Zimmermann*

Abstract

After decades of research, opinions are still split as to whether geminates should be represented as long or as heavy. In this paper, we attempt to resolve this riddle by entertaining a model that basically rests on the principle that while all geminates are underlyingly moraic, they might not emerge as such on the surface. This intuition – due to Davis (2011) – is formalized neatly through an extension of Containment Theory within OT (Prince and Smolensky 1993/2004) along the lines of Zimmermann (2014). We discuss the main typology of gemination (medially, initially, finally), briefly explore some predictions the system makes and illustrate how distinct patterns of gemination within the same language, e.g. weightful medial and final geminates, vs. weightless initial ones, as in Swiss German, can be generated. Importantly, in all cases investigated, we show how our model consistently manages to maintain a representational contrast between singletons and geminates.

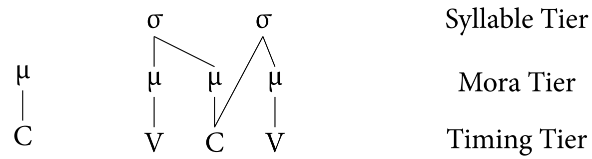
1. Introduction

A recurring, and yet unresolved, debate in the literature concerns the representation of geminates. Using the terminology of Ringen and Vago (2011) and Davis (2011), the two main competing theories are the syllabic weight analysis (1a) and the segmental length analysis (1b) of geminates.

(1) *Geminate representations* (Ringen and Vago 2011: 156)

a. The syllabic weight analysis of geminates

Underlying *Intervocalic*

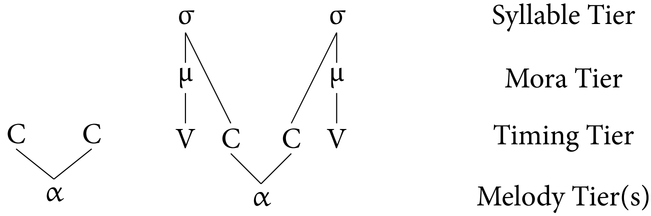


*We would like to thank Peter Staroverov for helpful comments on a previous version of this paper.

Topics at InfL, 65–90
A. Assmann, S. Bank, D. Georgi, T. Klein, P. Weisser & E. Zimmermann (eds.)
Linguistische Arbeits Berichte 92, Universität Leipzig 2014

b. The segmental length analysis of geminates

Underlying *Intervocalic*



In the syllabic weight analysis (Hayes 1989, Davis 1994, 1999*a*, 2003, Topintzi 2008, 2010), a geminate is underlyingly moraic. The moraic consonant is associated to the coda of one syllable and also to the onset of the next one. This second link ensures that the structure avoids an onsetless syllable. The fact then that the geminate typically straddles syllable boundaries intervocalically, producing what is known as a 'flopped' structure, is actually a product of syllabification considerations (Ham 2001). The splitting of the geminate also serves well to reflect the fact that geminates are typically longer, in fact up to three times longer (Ladefoged and Maddieson 1996: 91+92), than singletons. This property of geminates is what Ringen and Vago (2011) strive to capture in (1b). For them, weight – as in the intervocalic case here in (1b) – may emerge on the surface because of the constraint WeightByPosition, but does not have to. To sum up, under (1a) geminates are underlyingly heavy; under (1b), they are underlyingly long. For convenience, we will refer to those as the 'weight' vs. 'length' theories.

The two theories make very different predictions, many of which are thoroughly examined in Davis (2011). We will mention the major ones briefly here. In the 'length' theory, geminates project two C elements on the timing tier, therefore they should pattern like a string of two consonants when it comes to processes that refer to the C/V-tier. This is not the case for the 'weight' approach. Ringen and Vago (2011) claim that this prediction is borne out in Hungarian, a language that bans strings of 3 Cs and resolves such cases through epenthesis, thus CCC→CCεC. As anticipated in their framework, geminates produce the same effect, thus: /fygg-s/→[fygges] 'depend-2sg'. But geminates don't behave uniformly like that. In Hadhrami Arabic for example (Davis 2011), word-final clusters are split through epenthesis, e.g. /gird/→[girid] 'a monkey', word-final geminates are not, e.g. [rabb] 'Lord'. This asymmetry is not anticipated under (1b).

A strong argument in favor of the 'weight' theory is the following. Given that under (1b), weight – if applicable – is a derived property of geminates, it seems that geminates in coda position[1] should always pattern uniformly alongside the singleton codas in the language in either being both moraic or not (but not a mixture of the two). This is what Tranel (1991) originally called the Principle of Equal Weight for Codas. Such Principle predicts the patterns in (2I+II), but not those in (2III+IV). The 'weight' approach predicts instead (2II+III).

(2) *Weight for singleton (C) and geminate (G) codas*

	CVC	CVG
I.	light	light
II.	heavy	heavy
III.	light	heavy
IV.	heavy	light

There are arguments in the literature that all these possibilities are in fact attested. (2I) is represented by Selkup, Malayalam, Tübatulabal (Tranel 1991) or Chuvash (Davis 2011). Davis (2003, 2011: 890) however claims that possible re-analyses of the facts are available in order to maintain the 'weight' approach of geminates. (2II) is uncontroversial. Prime examples are Latin and Lake Miwok (Tranel 1991). Where things get to be really interesting are the cases in (2III-IV). In a number of papers, Davis (see 2011 and references therein), has convincingly argued for the existence of type (2III) languages. Among them are to be found: West Swedish, Koya, Seto, Fula, Cahuilla or Hausa. To our knowledge, these data have not been reanalyzed by the proponents of the 'length' approach. Finally, (2IV) is arguably illustrated by Ngalakgan. According to Baker (1998, 2008), codas in heterorganic clusters act as moraic for stress purposes, but codas in homorganic contexts, i.e. geminates or NC clusters, do not. We argue below that this pattern can fall out as well in a containment-based OT system where all geminates are underlyingly moraic.

The aim of this short survey has been to demonstrate that the discussion on geminate representation remains as alive as ever. While it has been generally claimed that most cases reported as evidence for the 'length' approach are

[1] Meaning consonants that are doubly linked to a coda and onset position. We term these 'coda geminate' in the following and abbreviate them with 'G'.

subject to reanalysis couched within the 'weight' approach, there is admittedly at least a handful of cases that are more straightforwardly captured under (1b), e.g. Hungarian epenthesis or Selkup and Ngalagkan stress.

Davis (2011) himself acknowledges this issue too and suggests various solutions, among which is the possibility that geminates may exhibit distinct representations on a language specific basis. However a more interesting solution is one that both Davis (2011: 893) and Ringen and Vago (2011: 166) foster, namely a unique representation of all geminates. Having considered a wealth of data, Davis envisages that such goal could be accomplished either through enrichment of representations – so that both a segmental and prosodic tiers are included, e.g. the Composite Model of Curtis (2003) – or by maintaining the underlying-mora approach of geminates with the added proviso that on the surface the structure may be altered.

In the present paper, we maintain the hypothesis that geminates are underlyingly moraic, but also entertain a proposal that blends these two suggestions in a novel way. To this end, we employ a version of containment in OT (Prince and Smolensky 1993/2004) as revitalized and further developed in Zimmermann (2014) (cf. also Trommer (2011) or Trommer and Zimmermann (to appear)). Our account provides further support for Zimmermann's account that was originally proposed to capture morphologically-triggered phonological length alternations. Conversely, from the viewpoint of gemination-theory, we will now be able to capture Davis' intuition – in relation to Ngalakgan – that "while geminates may be underlyingly moraic, they do not surface as moraic" (2011: 892) using extant and independently motivated machinery.

The remainder of the paper is structured as follows: In section 2, we present our theoretical background assumptions of containment-based OT and show how our system predicts four types of languages differing in whether coda singletons and geminates contribute to syllable weight. In section 3, we discuss the issue of edge geminates and their representation in our model. Brief case studies of initial geminates in Thurgovian Swiss and final geminates in Hungarian further illustrate the representational contrasts between singletons and geminates possible in a containment-based model. The factorial typology of the constraint system we adopt is briefly discussed in section 4. Relying on the software OT-Help, we argue that only attested grammars are predicted by our theory that differ in which types of geminates exist in a language, whether they contribute to syllable weight, and whether singletons contribute to syllable weight. We conclude in section 5.

2. The theory

2.1. Background assumptions: containment-based OT

Crucial to our present purpose is the assumption of containment that was already present in the discussion of OT in Prince and Smolensky (1993/2004) but was rejected in favor of a correspondence-theoretic OT (McCarthy and Prince 1995). Its original formulation given in (3) demands that the input is *contained* in the output and hence no literal deletion of structure is possible. Elements can only lack a phonetic interpretation if they are not integrated under the highest prosodic node and are consequently phonetically 'invisible'.

(3) *Containment* (Prince and Smolensky 1993/2004)
Every element of the phonological input representation is contained in the output.

Containment as we assume it here goes even further and demands that the input must be reconstructable from the output at any time. Consequently, not only elements like segments or features are subject to containment but association lines are as well and can never be deleted, only remain phonetically 'invisible' if they are marked as uninterpretable for the phonetics (Goldrick 2000, van Oostendorp 2006, Revithiadou 2007, Trommer 2011, Zimmermann 2014, Trommer and Zimmermann to appear). Consequently, there are four possible types of association lines, given in (4). Association lines can be underlyingly present and phonetically visible in the output (4a) (=notated as straight lines), they can be underlyingly present and marked as phonetically invisible (4b) (=notated as lines that are crossed out), they can be epenthetic and phonetically visible (4c) (=notated as dotted lines), or they can be epenthetic and phonetically invisible (4d) (=notated as dotted lines that are crossed out).[2]

[2] This typology of association lines is highly reminiscent of the system proposed in Turbidity Theory (Goldrick 2000, van Oostendorp 2006, Revithiadou 2007) where association lines are replaced with the two relations of projection and pronunciation. In Goldrick's original proposal of Turbidity Theory, the former denotes an abstract relationship between two elements and the latter denotes the output relations that are visible for the phonetics (Goldrick 2000).

(4) *Marking conventions for different types of association lines*

Morphological association lines		Epenthetic association lines	
phonetically visible:	phonetically invisible:	phonetically visible:	phonetically invisible:
a. │	b. ǂ	c. ⋮	d. ⋮̶

Under our containment model, phonological elements like segments, prosodic nodes, features, or tones are taken to be phonetically invisible if they are not properly integrated under the highest prosodic word node (Prince and Smolensky 1993/2004: 25). Since association lines can be marked as phonetically invisible, it is clear that a 'properly integrated' element that is visible for the phonetics implies that it is integrated under the highest prosodic node via phonetically visible association lines. Association lines, on the other hand, can only be phonetically visible if they associate a lower element to a phonetically visible higher one. This principle of phonetic visibility is summarized through the two statements in (5).

(5) *Principles of phonetic (in)visibility* (Zimmermann 2014: 49)

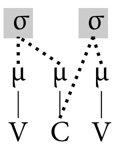
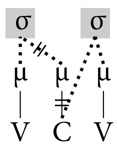
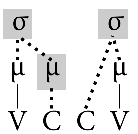
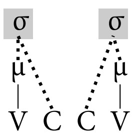
a. Every association line linking a phonetically invisible element to a lower element is *phonetically invisible*.

b. Every element is *phonetically invisible* iff it is not associated with a higher prosodic node by a phonetically visible association line.

A central assumption in this containment-based model is that constraints exist in at least two versions: one referring only to phonetically visible structure and another referring to *all* kinds of structure, including phonetically invisible one. This is explicitly formulated as the 'cloning hypothesis' in Trommer (2011). Another assumption that is necesssary in containment is the often implicit assumption that morphological affiliation is detectable. In van Oostendorp (2006) and Revithiadou (2007), this theoretical assumption is made explicit as the assumption of morphological colours: every morpheme has its own index (= 'colour') and all elements that are part of the underlyingly representation of this morpheme bear this index. This allows to distinguish whether two elements belong to the same morpheme or not and whether an element is epenthetic (hence colourless). Epenthetic elements are marked with a grey background in all following depictions.

These background assumptions now allow to predict all the attested types of languages in (2) while maintaining that geminates are distinguished from non-geminates by being underlyingly moraic. In languages where geminates do not contribute to syllable weight, the underlying association to the μ is simply marked as uninterpreted. It remains, however, in the structure and constraints sensitive to phonetically invisible structure can always refer to this structural difference between geminates and non-geminate consonants. Geminates and singletons can hence both be predicted to contribute to syllable weight (=phonetically associated to a μ) or to be irrelevant for syllable weight (=not phonetically associated to a μ). This is briefly summarized in (6): the underlying association of a consonant to its μ can be phonetically visible (6a) or not (6b) and an underlyingly μ-less consonant might remain μ-less (6d) or might project an epenthetic μ (6c). Crucially, as we show in detail in section 2.2, for this choice of being moraic or not, different constraints are relevant for singletons and geminates. Hence, the (non)moraicity of geminates is not bound to the (non)moraicity of singleton codas in our model – in contrast to the prediction of the principle of 'Equal Weight for Codas' (Tranel 1991).

(6) *Underlyingly (non)moraic consonants and syllable weight*

	…can contribute to syllable weight	**…can be irrelevant for syllable weight**
Geminate: μ | C	a. 	b. 
Non-geminate: C	c. 	d. 

Of special interest here is the representation we assume in (6b), namely the representation of a geminate that does not contribute to syllable weight. This results from the fact that the association of the consonant-μ to the syllable is marked as phonetically invisible and is hence not interpreted by the phonetics.

In containment theory, however, this underlying μ remains in the structure and ensures that the consonant is still doubly linked to two syllables – the structural difference that distinguishes it from a non-geminate consonant.[3]

2.2. Predicting the four language types

In the following subsection, we illustrate how the four types of languages (7) (repeated from (2)) that differ in the question of whether singleton and/or geminate codas contribute to syllable weight fall out in our containment-based system assuming that all geminates are underlyingly moraic.

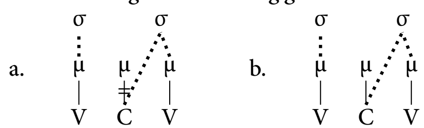
(7) *Weight for singleton (C) and geminate (G) codas*

	CVC	CVG	*example*
I.	light	light	*Selkup*
II.	heavy	heavy	*Latin*
III.	light	heavy	*Hausa*
IV.	heavy	light	*Ngalakgan*

The relevant constraints are given in (8). ONS, WBP, and $*C^{\mu}$ are standard markedness constraints on syllable structure, discussed in, for example, Prince and Smolensky (1993/2004) or Sherer (1994). $*_{\sigma}[C^{\mu}$ is a specific version of $*C^{\mu}$ that bans moraic onsets – a necessary markedness constraint given the assumption that moraic onset are highly marked but possible (Topintzi 2008, 2010). The existence of MAX(μ—S) is a necessary consequence of the containment-based system we adopt: it penalizes association lines that are

[3]There are other imaginable structures that are in principle possible in a containment-based system for such a non-weight bearing geminate that we cannot discuss in detail for reasons of space. That the phonetically invisible μ of a geminate is still integrated under a syllable node follows from a high-ranked demand that every μ must be integrated under a syllable. However, in principle, a structure (ia) is also possible. (ib) is generally excluded in our system that assumes the principles (5) since a phonetically invisible μ dominates a C in a phonetically visible way. However, we leave it open for future research whether those structures might receive independent support from other patterns of weight distribution.

(i) *Other non-weight contributing geminates*



marked as invisible and hence mirrors roughly the effect of a faithfulness constraint demanding the preservation of underlying prosodic structure.

(8) a. ONSET (=ONS)
Assign a violation mark for every syllable without an onset consonant.

b. WEIGHTBYPOSITION (=WBP)
Assign a violation mark for every coda consonant that is not phonetically dominated by a μ.

c. $*C^{\mu}$
Assign a violation mark for every consonant that is phonetically dominated by a μ.

d. $*_{\sigma}[C^{\mu}$
Assign a violation mark for every onset consonant that is phonetically dominated by a μ.

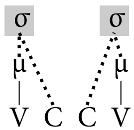
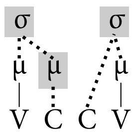
e. MAX(μ—S)
Assign a violation mark for every phonetically invisible association line between a μ and a segment.

In a first possible ranking of these constraints, high-ranked $*C^{\mu}$ ensures that no consonant is associated to a μ in a phonetically visible way. Coda consonants hence never become moraic (9i) and the underlying association line from a geminate consonant to its μ is marked as phonetically invisible (9ii). The non-moraic singleton coda in candidate (9i-a) induces a violation of WBP but since $*C^{\mu}$ is high-ranked, this candidate becomes optimal. The non-realization of an underlyingly present μ dominating a consonant (9ii-b), on the other hand, induces a violation of MAX(μ—S) but since this constraint is dominated by $*C^{\mu}$ as well, this candidate wins the competition. It does not violate WBP since the consonant is not phonetically visibly linked to a coda position and is hence irrelevant for calculating violations of WBP. Note that a candidate that leaves the invisible μ unintegrated under a syllable node is not listed in (9ii). Such a structure would not make any different prediction with respect to the phonetic interpretation and we assume that such a candidate is excluded by a high-ranked constraint ensuring that every μ must be dominated by a syllable (cf. our discussion in footnote 3). The sub-optimal candidates (9ii-a), (9ii-c), and (9ii-d) show different strategies to integrate the underlying consonant μ in a phonetically visible way: the consonant can be integrated under two syllables

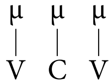
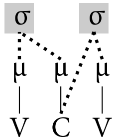
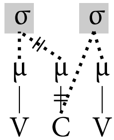
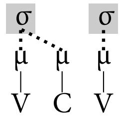
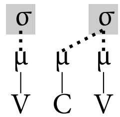
(9ii-a), it can be integrated only as coda additionally violating Ons (9ii-c), or it can be integrated as moraic onset additionally violating $*_{\sigma}[C^{\mu}$ (9ii-d). For WbP and $*_{\sigma}[C^{\mu}$, only phonetically visible associations of consonants to the syllable positions coda or onset are relevant. (9ii-a) hence does not violate $*_{\sigma}[C^{\mu}$ although the consonant is syllabified as onset and is indeed moraic – however, the link to the onset position is direct and does not include a μ. Recall that epenthetic elements are marked with a grey background.

(9) *Type I: Selkup*

i. *CVC=light*

	$*C^{\mu}$	$*_{\sigma}[C^{\mu}$	Max μ—s	WbP	Ons
☞ a. 				*	
b. 	*!				

ii. *CVG=light*

	$^*C^{\mu}$	$^*{}_{\sigma}[C^{\mu}$	MAX μ—S	WBP	ONS
a. 	*!				
☞ b. 			*		
c. 	*!				*
d. 	*!	*			

This ranking predicts the type I language we discussed in (7), exemplified by, for example, Selkup. The second type of language that can be predicted is a language where all codas contribute to syllable weight, irrespective of whether they are singleton consonants or the first part of a geminate (type II in (7)). That geminate consonants always contribute to syllable weight follows if MAX(μ—S) is ranked above $^*C^{\mu}$ and hence the underlying μ of a geminate consonant is always phonetically realized, even if this implies a violation of $^*C^{\mu}$. And if WBP outranks $^*C^{\mu}$ as well, every singleton coda will project an epenthetic μ. This ranking and its outcome is summarized in (10). Note that in the following tableaux, we abbreviate the full autosegmental structures we gave in (9) and notate a segment that is associated with a μ in a phonetically visible way with X^{μ} and a segment that is associated to a μ through an underlying but phonetically invisible association line by $X^{(\mu)}$. For ease of exposition, we give the same candidates as in (9) in the same order in all the following tableaux.

(10) *Type II: Latin*

	WbP	Max μ—S	$*C^{\mu}$	$*_{\sigma}[C^{\mu}$	Ons
$V^{\mu}CCV^{\mu}$	i. CVC=heavy				
a. $V^{\mu}CCV^{\mu}$	*!				
☞ b. $V^{\mu}C^{\mu}CV^{\mu}$			*		
$V^{\mu}C^{\mu}V^{\mu}$	ii. CVG=heavy				
☞ a. $V^{\mu}C^{\mu}V^{\mu}$			*		
b. $V^{\mu}C^{(\mu)}V^{\mu}$		*!			
c. $V^{\mu}C^{\mu}V^{\mu}$			*		*!
d. $V^{\mu}C^{\mu}V^{\mu}$			*	*!	

If now Max(μ—S) is ranked above $*C^{\mu}$ but WbP is ranked below $*C^{\mu}$, we expect a pattern where only underlyingly moraic consonants surface as moraic but no epenthetic μ is inserted for an underlyingly μ-less coda. The effect of this ranking is shown in tableaux (11) that predict the type III language we listed in (7). This is the first pattern that does not follow the Equal Coda Weight Principle but is, as we mentioned above, attested in, for example, Hausa.

(11) *Type III: Hausa*

	$*_{\sigma}[C^{\mu}$	Max μ—S	Ons	$*C^{\mu}$	WbP
$V^{\mu}CCV^{\mu}$	i. CVC=light				
☞ a. $V^{\mu}C\ CV^{\mu}$					*
b. $V^{\mu}C^{\mu}CV^{\mu}$				*!	
$V^{\mu}C^{\mu}V^{\mu}$	ii. CVG=heavy				
☞ a. $V^{\mu}C^{\mu}V^{\mu}$				*	
b. $V^{\mu}C^{(\mu)}V^{\mu}$		*!			
c. $V^{\mu}C^{\mu}V^{\mu}$			*!	*	
d. $V^{\mu}C^{\mu}V^{\mu}$	*!			*	

And finally, if WbP dominates $*C^{\mu}$ that in turn dominates Max(μ—S), we expect a pattern where singleton codas contribute to syllable weight but underlyingly moraic intervocalic consonants do not. This ranking given in (12) predicts a type IV language that is exemplified with Ngalakgan (7). Undominated

WBP demands insertion of epenthetic μ's for coda consonants (12i) but the underlying μ of a geminate remains phonetically unrealized (12ii). This follows since the underlyingly moraic consonant can be realized as a mora-less onset: a violation of $*C^{\mu}$ can be avoided without inducing a new violation of WBP. MAX(μ—S) demanding preservation of the underlying μ is too low-ranked to have an effect.

(12) *Type IV: Ngalakgan*

	WBP	ONS	$*_{\sigma}[C^{\mu}$	$*C^{\mu}$	MAX μ—S
$V^{\mu}CCV^{\mu}$	i. CVC=heavy				
a. $V^{\mu}CCV^{\mu}$	*!				
☞ b. $V^{\mu}C^{\mu}CV^{\mu}$				*	
$V^{\mu}C^{\mu}V^{\mu}$	ii. CVG=light				
a. $V^{\mu}C^{\mu}V^{\mu}$				*!	
☞ b. $V^{\mu}C^{(\mu)}V^{\mu}$					*
c. $V^{\mu}C^{\mu}V^{\mu}$		*!		*	
d. $V^{\mu}C^{\mu}V^{\mu}$			*!	*	

This concludes our illustration how the general language types in (2) follow in a containment-based OT model under the assumption that geminates are underlyingly moraic. All four patterns differing in whether geminates and codas contribute to syllable weight can be predicted in the model we propose.

3. Edge geminates

Generally, it is uncontested that intervocalic geminates are by far the most common crosslinguistically while geminates at word edges are far less frequent (Thurgood 1993, Muller 2001, Davis 2011, Dimitrieva 2012). Still, such geminates exist as well and in this section we discuss their representation in our model in some more detail. It will be shown that the typology in (7) was a simplification and there are languages where the question of whether geminates contribute to syllable weight depends on their position in the word. For reasons of space we cannot look into this question thoroughly but will focus on some interesting asymmetries found for initial geminates in subsection 3.1 and the general question of final geminates in subsection 3.2.

3.1. Initial geminates

In the ranking in (11) for the type III language Hausa, $*_{\sigma}[C^{\mu}$ dominates all faithfulness constraints, especially Max(μ—S). It is hence predicted that no moraic onset ever surfaces in this language and all underlying association lines between μ's and consonants that are syllabified as onsets are marked as phonetically invisible. There are, however, some examples of initial weight-bearing geminates. One language of this type is Trukese, an Austronesian language spoken in the Truk state of Micronesia (Hart 1991, Davis and Torretta 1998, Davis 1999*b*). All consonants in Trukese except the glides may surface as geminates and geminates are possible initially and medially. One piece of evidence for the surprising fact that initial geminates indeed contribute to weight comes from empirical facts on word minimality restrictions. Nouns must be CːV, CVː, or bisyllabic, but CV or CVC nouns are generally impossible. Trukese hence is a pattern III language where initial and medial geminates contributing to syllable weight. Such a pattern is predicted from a grammar that differs only slightly from the ranking we gave for Hausa in (11): if Max(μ—S) and $*_{\sigma}[C^{\mu}$ reverse their position in the ranking, it is predicted that initial geminates surface as long and weight-contributing. This can be seen in the additional context of an initial moraic consonant in (13iii). In Hausa, on the other hand, such an input would be neutralized to a non-moraic initial onset as can be seen in (14).

(13) *Type III: Trukese*

	MAX μ—S	$*_{\sigma}[C^{\mu}$	ONS	$*C^{\mu}$	WBP
$V^{\mu}CCV^{\mu}$	i. CVC=light				
☞ a. $V^{\mu}C\ CV^{\mu}$					*
b. $V^{\mu}C^{\mu}CV^{\mu}$				*!	
$V^{\mu}C^{\mu}V^{\mu}$	ii. CVG=heavy				
☞ a. $V^{\mu}C^{\mu}V^{\mu}$				*	
b. $V^{\mu}C^{(\mu)}V^{\mu}$		*!			
c. $V^{\mu}C^{\mu}V^{\mu}$			*!	*	
d. $V^{\mu}C^{\mu}V^{\mu}$	*!			*	
$C^{\mu}V$	iii. GV=heavy				
☞ a. $C^{\mu}V$		*			*
b. $C^{(\mu)}V$	*!			*	

(14) *Type III: Hausa, contd. from (11)*

	$*_{\sigma}[C^{\mu}$	MAX μ—S	ONS	$*C^{\mu}$	WBP
$C^{\mu}V$	iii. GV=light				
a. $C^{\mu}V$	*!			*	
☞ b. $C^{(\mu)}V$		*			

Thurgovian Swiss, a Swiss dialect spoken in the canton Thurgau (Muller 2001, Kraehenmann 2001, 2003), now shows an alternating behaviour of geminates with respect to their weight contribution. The language has a length contrast for vowels and consonants and geminates surface in all positions. Examples for medial (15a), final (15b), and initial (15c) geminates are given below, contrasted with singleton (near) minimal pairs.

(15) *Geminates in Thurgovian Swiss* (Kraehenmann 2003: 42+43)

	Geminates		Singletons	
a.	matːə	'mat'	matə	'maggot'
	jakːə	'jacket'	jakə	'to hunt'
b.	hilfː	'help'	ʃilf	'reed'
	ʀuːʃː	'inebriation'	ʀuːʃ	'rouge'
c.	pːuːtəʀ	'powder'	puːtə	'hut'
	tːaŋkː	'tank'	taŋkː	'thank'

There are no CV or CVC words in Thurgovian Swiss. Instead words of the shape CVCC or CVG are allowed. This distributional fact easily follows under the assumption that a word minimality condition demands that words are minimally bimoraic in the language and codas are moraic but final consonants are extrametrical (Muller 2001). CVC words then undergo vowel lengthening in order to conform to this bimoraicity requirement, cf. the contrast in (16a). Words ending in a final geminate (16b) are not subject to vowel lengthening, implying that the geminate contributes a μ to syllable weight.[4] Interestingly now, words with an initial geminate do undergo vowel lengthening as well (16c) and no word GV(C) ever surfaces. One can hence conclude that medial and final geminates are weight-contributing in Thurgovian Swiss, initial ones are not.

(16) *Word minimality in Thurgovian Swiss* (Muller 2001: 101)[5]

	Root	Singular	Plural	
a.	/has/	haːs	hase	'hare'
b.	/fɛtː/	fɛtː	fɛtːe	'fat'
c.	/tːak/	tːaːk	tːake	'day'

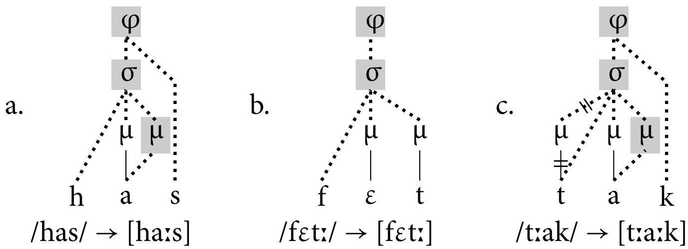
(17) lists the structures in our containment-based model that derive this asymmetric behaviour of geminates. (17a) shows a simple CVC stem: since the final C is extrametrical, an additional μ resulting in vowel lengthening is required to conform to the word minimality requirement. Medial and final geminates (17b) remain phonetically associated to the μ they were underlyingly associated with and no additional vowel lengthening to ensure bimoraicity is required.

[4] Underlyingly moraic consonants are exempt from the final extrametricality. This is easily derivable if μ's prefer not to be associated directly to a foot node (=DepAL$_{\mu-\varphi}$).

[5] Note that Muller (2001) does not provide the data in consistent IPA notation.

Initial geminates (17d), however, cannot remain phonetically associated to their μ due to high-ranked $*_{\sigma}[C^{\mu}$ and additional vowel lengthening is hence required.

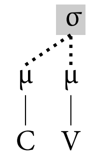
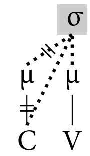
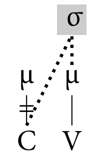
(17) *Word minimality and vowel lengthening in Thurgovian Swiss*



Initial geminates in Thurgovian Swiss are hence not phonetically dominated by a μ but since the underlying μ always remains in the structure in containment, the geminate remains structurally different from a singleton onset: it is doubly linked to a syllable (even though one of the association paths is phonetically invisible). And this double linking is then interpreted as length. Note that the consonant would remain phonetically unrealized if it were only integrated via the phonetically invisible association path to the μ; the additional epenthetic association line is hence demanded by high-ranked Max-S ensuring that no segment is deleted.

To summarize this very brief typology: Hausa has no initial geminates but medial ones that contribute to syllable weight, Trukese has medial and initial geminates that both contribute to syllable weight, and Thurgovian Swiss has medial and initial geminates but only the former contribute to syllable weight. The representational contrast between Trukese and Thurgovian Swiss, we argue, is that the underlying μ of an initial consonant is phonetically visibly integrated in the former (18a) but not in the latter (18b). Interestingly, given the assumption of Richness of the Base in OT, languages like Hausa without initial geminates must still deal with a possible underlying representation of a moraic initial consonant. We assume that in such a language, the representation in (18c) predicts that the initial consonant neither contributes to syllable weight nor is realized as long – the μ is not integrated into the syllable structure and the consonant not doubly linked; not even through a phonetically invisible association line.

(18) *Initial moraic consonants*

	Trukese	T. Swiss	Hausa
	a. 	b. 	c. 
phonetics:	[Cː V]	[Cː V]	[CV]
contributes to syllable weight:	yes	no	no

3.2. Final geminates

An apparent conundrum for the moraic approach to geminates are languages where geminates and codas are moraic and where final geminates exist. In Ringen and Vago (2006), it has been argued that Hungarian is such a language and contrasts as in (19) for final singletons and geminates exist. There is a word minimality restriction active in the language that excludes CV words. This follows if words are minimally bimoraic in Hungarian and both codas and geminates contribute to syllable weight.

(19) *Final singletons vs. geminates* (Ringen and Vago 2006: 14)

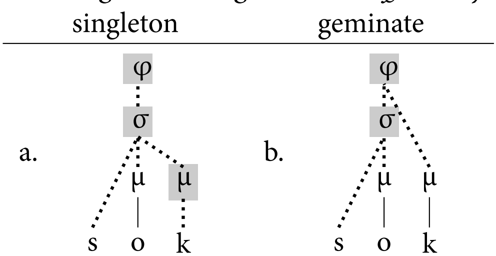
	singleton			geminate	
a.	sok	'many'	b.	sokː	'shock'
	hal	'fish'		halː	'hear(s)'
	tol	'push(es)'		tolː	'pen'

For a moraic approach, this is problematic at first sight: what distinguishes the singletons in (19a) from the geminates in (19b), the latter being underlyingly moraic, the former receiving a μ due to WbP? Notice, however, that other interpretations of the empirical facts are available, such as the non-weight-contributing role of singleton codas (cf., for example, Grimes 2010). But even if the Hungarian facts are as described in Ringen and Vago (2011), our approach straightforwardly offers a solution since the distinction into underlying and inserted structure is always detectable in containment. Recall the assumption we introduced in section 2.1 that the morphological affiliation of every phonological element is visible and epenthetic elements can be identified since they are lack any morphological colour. The underlying μ's in (19b) can hence be distinguished from the μ's in (19a) since the former bear a morphological colour, the latter do not. Because constraints such as DEPAL can be sensitive to this distinction, it is predicted that colourless and morphologically coloured μ's can be syllabified differently. More concretely, the constraint in (20) demanding that no association line between an underlying μ and a syllable should be inserted, prohibits an integration of an underlying μinto a syllable.

(20) DEPAL$_{\sigma\text{-}\mu}$
Assign a violation mark for every colourless association line between a σ and a morphologically coloured μ.

If (20) is high-ranked, a structure as in (21b) is predicted where an underlying μ is extrasyllabic and directly attached to a foot node. An epenthetic μ, on the other hand, can easily be integrated under a syllable node (21a).[6]

(21) *Final singletons and geminates: different syllabification structure*



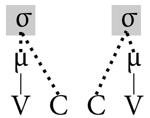
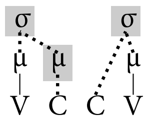
4. A full typology: geminates and syllable weight

A remaining question is whether the phonological structures we assume together with the constraint system we employ predict any unattested combinations of (non)weight contributing geminates and singletons. In this section we discuss the factorial typology based on the OT-Help software and argue that all grammars predicted are indeed attested. For reasons of space, we limit our discussion to four relevant contexts (singleton codas, medial geminates, initial geminates, and final geminates) and only include the most relevant candidates

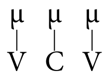
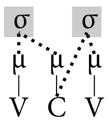
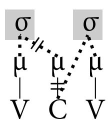
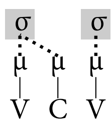
[6] Cf. Trommer and Zimmermann (to appear) for a more detailed discussion of this implementation of DEPAL in containment in the domain of morphological μ affixation.

and constraints.[7] Tableaux (22) to (25) give the set of relevant candidates and list their constraint violations. The constraints are the same we introduced in (8) and are not ranked with respect to each other. Tableaux (24) and (25) are especially interesting since they summarize the structures discussed in section 3 for edge geminates.

(22) *Non-moraic medial consonant*

	*C^{μ}	*$_{\sigma}[C^{\mu}$	MAX μ—S	WBP	ONS
a. 				*	
b. 	*				

(23) *Medial geminate*

	*C^{μ}	*$_{\sigma}[C^{\mu}$	MAX μ—S	WBP	ONS
a. 	*				
b. 			*		
c. 	*				*

[7] An example for an issue that we need to exclude in the following is the distinction between medial and final singleton codas and hence the effect of NONFINALITY (Hyde 2011).

(24) *Initial geminate*

μ μ | | C V	$*C^{\mu}$	$*_{\sigma}[C^{\mu}$	MAX μ—S	WBP	ONS
a. σ; μ μ; C V	*	*			
b. σ; μ μ; C V			*		

(25) *Final geminate*

μ μ | | V C	$*C^{\mu}$	$*_{\sigma}[C^{\mu}$	MAX μ—S	WBP	ONS
a. σ; μ μ; V C	*				
b. σ; μ μ; V C			*	*	

We fed these 4 tableaux with a total of 9 candidates into the software OT-Help (Staubs et al. 2010) that automatically calculated the possible grammars. OT-Help found 6 possible parallel OT-grammars, listed below in (26). For all of the 4 relevant contexts (22) to (25), we abbreviate the winning candidates for every predicted language. For non-moraic medial consonants (22), either the non-moraic consonant (a.) or the consonant dominated by an epenthetic μ (b.) wins; for medial moraic consonants, either the structure with a consonant that is phonetically visibly double-linked to two syllables (a.) or the structure where this double association is phonetically invisible (b) becomes optimal; and for initial (24) and final (25) moraic consonants, either a phonetically visible integration of the moraic consonant (a.) or a phonetically invisible double integration (b.) wins. All optimal candidates where a μ is phonetically visibly integrated are marked with a grey background to ease readability.

(26) *Predicted typology: 6 grammars*

Input:	Non-moraic C medial (22)	Moraic C medial (23)	initial (24)	final (25)
Lg1	non-moraic	vis.doubl	moraic	moraic
Lg2	moraic	vis.doubl	moraic	moraic
Lg3	non-moraic	vis.doubl	inv.doubl	moraic
Lg4	moraic	vis.doubl	inv.doubl	moraic
Lg5	moraic	inv.doubl	inv.doubl	moraic
Lg6	non-moraic	inv.doubl	inv.doubl	inv.doubl

We argue now that all these predicted rankings result in attested grammars. In (27), we list an example for each of these patterns. For every winning candidate, we indicate whether the consonant is phonetically visibly associated to a μ at the top of the cell and how this consonant is phonetically interpreted at the bottom of the cell. It became already clear in the preceding discussion in section 3 that the structures involving phonetically invisible structures can be interpreted differently. This concerns the b. candidates in (23) to (25): they are doubly associated but one of the associations is phonetically invisible. We argue that it is a language-specific choice whether such an association is interpreted as length or not. For (27-5), for example, we assume that the phonetically invisible double association of an onset to the same syllable is not interpreted as length, the phonetically invisible double association of a medial consonant to two different syllables, however, is indeed interpreted as length. This correctly predicts that there are no initial geminates in Ngalakgan but medial geminates that do not contribute to syllable weight.

Another asymmetry in the interpretation of phonological structure concerns moraic final consonants. In Latin (27-4), a final moraic consonant is simply interpreted as a singleton consonant whereas in Thurgovian Swiss (27-3), such a moraic final consonant is interpreted as geminate. Such asymmetries are simply due to different contrasts in a language: whereas Thurgovian Swiss exhibits non-moraic codas as well (=those that were underlyingly μ-less), all codas in Latin are moraic and hence interpreted the same. A final note is in order with respect to Marshallese that we cite as an example for language (27-1). At this point, the exclusion of some relevant constraints becomes problematic: in Marshallese, there are neither final geminates nor any moraic final consonants in general. This simply follows from the effect of well-established NonFinality (cf. our footnote 7). We hence notate it as a language 1'.

(27) *Attested typology: 6 grammars*

	type	N-mor.C (22)	Moraic C (23)	(24)	(25)	example
Lg1'	III	¬μ C	μ C:	μ C:	¬μ C	Marshallese (Topintzi 2008)
Lg2	II	μ C	μ C:	μ C:	μ C:	Tamazight Berber (Jebbour 1999, Kraehenmann 2011)
Lg3	III	¬μ C	μ C:	¬μ C:	μ C:	Thurgovian Swiss (Muller 2001, Kraehenmann 2001)
Lg4	II	μ C	μ C:	¬μ C	μ C	Latin (Mester 1994)
Lg5	IV	μ C	¬μ C:	¬μ C	μ C	Ngalakan (Davis 2011)
Lg6	I	¬μ C	¬μ C:	¬μ C	¬μ C	Selkup (Ringen and Vago 2011)

It has to be emphasized again, that this was only a preliminary typology that excluded several relevant constraints and contexts. It is hence clear that not all attested patterns of weight contribution and geminates can be captured with this set of constraints. Our aim was only to show that the constraints and structures we assumed to predict the four types of languages (7) are unproblematic from an empirical perspective and that all imaginable rankings of these constraint result in attested grammars.

5. Conclusion

To sum up, a core feature of our paper has been the assumption that what sets geminates apart from singletons is their underlying moraicity (Hayes 1989, Davis 1994, 1999*a*, 2003, Topintzi 2008, 2010). To account for geminates, however, that behave as weightless, we have further entertained Davis' (2011) idea that such underlying moraicity may not always survive on the surface. We have then developed a Containment based model, along the lines of Zimmermann (2014) that formalizes these basic principles. Our system correctly generates all main patterns. Importantly, we are also able to capture asymmetries with respect to geminates within a language. For example, we can produce languages such as Swiss German that exhibits geminates in all positions, but where medial and final ones are weightful whereas the initial ones are not. Moreover, we are able to always maintain a representational contrast between singletons and geminates, even when they are quite atypical, as in e.g. Ngalakgan where singletons behave moraically, but geminates do not. Our future plans involve conducting a fuller typology and exploring further the predictions our systems makes.

References

Baker, Brett (1998): Edge-Crispness: Segment to mora isomorphism. In: *Proceedings of the West Coast Conference on Formal Linguistics 17*. pp. 33–47.

Baker, Brett (2008): *Word Structure in Ngalakgan*. CSLI, Stanford.

Curtis, Emily (2003): Geminate weight: Case studies and formal models. PhD thesis, University of Washington.

Davis, Stuart (1994): Geminate consonants in moraic phonology. In: *Proceedings of WCCFL 13*. pp. 32–45.

Davis, Stuart (1999*a*): On the moraic representation of underlying geminates: Evidence from prosodic morphology. *In:* R. Kager, H. van der Hulst and W. Zonneveld, eds, *The prosody-morphology interface*. Cambridge University Press, pp. 39–61.

Davis, Stuart (1999*b*): 'On the Representation of Initial Geminates', *Phonology* **16**, 93–104.

Davis, Stuart (2003): The controversy over geminates and syllable weight. *In:* C. Féry and R. van de Vijver, eds, *The Syllable in Optimality Theory*. Cambridge University Press, pp. 77–98.

Davis, Stuart (2011): Geminates. *In:* M. van Oostendorp, C. J. Ewen, E. Hume and K. Rice, eds, *Companion to Phonology*. Wiley Blackwell, Malden MA, chapter 37.

Davis, Stuart and Gina Torretta (1998): An Optimality-Theoretic Account of Compensatory Lengthening and Geminate Throwback in Trukese. *In:* P. Tamanji and K. Kusumoto, eds, *Proceedings of NELS 28*. GLSA, pp. 111–125.

Dimitrieva, Olga (2012): Geminate typology and the perception of consonant duration. PhD thesis, Stanford University.

Goldrick, Matthew (2000): Turbid output representations and the unity of opacity. *In:* A. Coetzee, N. Hall and J. Kim, eds, *NELS 30*. GLSA, Amherst, MA, pp. 231–245.

Grimes, Stephen (2010): Quantitative investigations in Hungarian phonotactics and syllable structure. PhD thesis, Indiana University.

Ham, William (2001): *Phonetic and Phonological Aspects of Geminate Timing*. Routledge, New York, London.

Hart, Michele (1991): The moraic status of initial geminates in Trukese. In: *Proceedings of the 7th annual meeting of the Berkeley Linguistics Society*. pp. 107–120.

Hayes, Bruce (1989): 'Compensatory Lengthening in moraic phonology', *Linguistic Inquiry* 20, 253–306.

Hyde, Brett (2011): Extrametricality and Non-finality. *In:* M. van Oostendorp, C. J. Ewen, E. Hume and K. Rice, eds, *The Blackwell Companion to Phonology*. Blackwell, chapter 43.

Jebbour, Ahmed (1999): 'Syllable weight and syllable nuclei in Tashlhiyt Berber of Tiznit', *Cahiers de Grammaire* 24, 95–116.

Kraehenmann, Astrid (2001): 'Swiss German Stops: Geminates All over the Word', *Phonology* 18, 109–145.

Kraehenmann, Astrid (2003): *Quantity and Prosodic Asymmetries in Alemannic: Synchronic and Diachronic Perspectives*. Mouton de Gruyter, Berlin, New York.

Kraehenmann, Astrid (2011): Initial geminates. *In:* M. van Oostendorp, C. J. Ewen, E. Hume and K. Rice, eds, *The Blackwell Companion to Phonology*. Wiley Blackwell, Malden MA, chapter 47.

Ladefoged, Peter and Ian Maddieson (1996): *The Sounds of the World's Languages*. Blackwell.

McCarthy, John and Alan Prince (1995): Faithfulness and Reduplicative Identity. *In:* J. Beckman, L. Dickey and S. Urbanczyk, eds, *University of Massachusetts Occasional Papers in Linguistics*. GLSA, Amherst, MA, pp. 249–384.

Mester, Armin (1994): 'The quantitative trochee in Latin', *NLLT* 12, 1–61.

Muller, Jennifer (2001): The Phonology and Phonetics of Word-initial Geminates. PhD thesis, Ohio State University.

Prince, Alan and Paul Smolensky (1993/2004): *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell, [first circulated as Prince & Smolensky (1993) Technical reports of the Rutgers University Center of Cognitive Science].

Revithiadou, Anthi (2007): Colored turbid accents and containment: A case study from lexical stress. *In:* S. Blaho, P. Bye and M. Krämer, eds, *Freedom of Analysis?*. Mouton De Gruyter, Berlin, New York, pp. 149–174.

Ringen, Catherine and Robert Vago (2006): 'Geminates: Heavy or long?', Ms., prepublished version.

Ringen, Catherine and Robert Vago (2011): Geminates: Heavy or long?. *In:* C. Cairns and E. Raimy, eds, *Handbook of the syllable*. Brill, Leiden, pp. 155–169.

Sherer, Tim D. (1994): Prosodic Phonotactics. PhD thesis, University of Massachusetts at Amherst.

Staubs, Robert, Michael Becker, Christopher Potts, Patrick Pratt, John McCarthy and

Joe Pater (2010): 'OT-Help 2.0. Software package.', Amherst, MA: University of Massachusetts Amherst.

Thurgood, Graham (1993): Geminates: A Cross-linguistic Examination. *In:* J. A. Nevis, G. McMenamin and G. Thurgood, eds, *Papers in Honor of Frederick H. Brengelman on the Occasion of the Twenty-Fifth Anniversary of the Department of Linguistics, CSU Fresno*. California State University Fresno, pp. 129–139.

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, Cambridge.

Tranel, Bernard (1991): 'CVC light syllables, geminates and moraic theory', *Phonology* **8**, 291–302.

Trommer, Jochen (2011): 'Phonological aspects of Western Nilotic Mutation Morphology', Habil. University of Leipzig.

Trommer, Jochen and Eva Zimmermann (to appear): 'Generalized Mora Affixation', *Phonology* .

van Oostendorp, Marc (2006): 'A Theory of Morphosyntactic Colours', Ms., Meertens Institute, Amsterdam, available online at http://egg.auf.net/06/docs/Hdt

Zimmermann, Eva (2014): A phonological account of morphological length. PhD thesis, Leipzig University.