# Geminate Consonants: A Literature Review

Lavinia Dunagan

April 2019

## Contents

# 1 Introduction

The sea changes which have occurred in linguistics in the past 50 years have motivated the re-analysis of a variety of phonological phenomena. One of these phenomena is geminate (or long) consonants.[1] These consonants, distinguished by their duration and sometimes double release from "singleton" counterparts, exhibit a number of qualities highly relevant to moraic theory, autosegmental representations, and broadly a more comprehensive account of phonemic contrasts. Geminates operate on multiple levels of linguistic inquiry, and so offer both problems and important counterpoints for standard phonological models (Curtis 2003).

In the history of linguistics there have been several reevaluations of geminates, though the most significant appear to have passed. Structuralist descriptions generally declined to consider phonemic geminates, instead assuming that length contrasts occur on a morphophonemic level.[2] In the early days of generative phonology, geminates were largely regarded as the [+long] versions of their singleton counterparts. That model has been rejected, since the feature representation does not satisfactorily account for the lack of rules using [+long] and the occasions when geminates pattern like two consonants (Leben 1980; Curtis 2003). Virtually every current account of geminates relies on an autosegmental model, and usually involves some moraic feature. Within this paradigm there are various accounts of how weight is attached to geminates, and some controversy over what characteristics are universal and what are linguistically contingent.

[1] In this paper "geminate" refers exclusively to underlying geminate consonants.

[2] It bears noting that geminates do behave differently from sequences of two consonants. This contrast will be discussed more in the "Unit Integrity" section.

## 1.1 A Note on Phonetics and Perception

Geminates are principally characterized by duration, although secondary cues are fairly common (Swadesh 1937; Muller 2001). The singleton/geminate distinction is most simply described as a difference in onset (for stops and usually affricates) or maintenance (for fricatives); the existence of a geminate implies a singleton counterpart (Davis 2011). This contrast is distinctly dependent on the environment of the geminate, and so bears examination as a possible contributing factor to some of the stranger aspects of geminate behavior (Muller 2001). The perceptual characteristics of geminates likely influence their phonotactic treatment. While one does not encounter word-initial geminates in many languages, the work done on them by Jennifer Muller and others suggests some elements of the phonology-phonetics interface might be evinced by geminates (Muller 2001). In Muller's study of Chuukese, the secondary cues for geminates are very subtle; however, word-initial geminates are still "demonstrably moraic," i.e. distinct from singletons, and distinct from consonant clusters - which do not appear in Chuukese (Muller 2001). Generally, geminates are influenced by their environments on a cross-linguistic level. We might regard medial geminates as the most common type; even so, geminates definitely occur word-initially and word-finally (Muller 2001; Davis 2011).

## 1.2 A Note on Typology

Geminates are hardly alone in the phonological literature as a feature whose commonly discussed data is largely limited to a couple languages; even so, some quite well-regarded accounts bear the mark of narrowness. As is apparent in Davis's review, if one looks at all the studies of geminate consonants produced, there are many (sometimes consistent, sometimes not) ways of analyzing

them (Davis 2011). While geminates do exhibit some cross-linguistic properties, they appear to be highly changeable based on the language. Emily Curtis notes that all of the popular models of geminates in relation to syllables do not correctly model every language (Curtis 2003).

In the first place, geminate consonants are uncommon. According to UPSID, the single geminate consonant which occurs the most, [sː], appears in 1% of the sampled languages (Maddieson and Precoda 1989). While this means that 1% is only the minimum number of languages with geminates, it still suggests that the number is quite small. In light of this history, the tradition of geminate phonology is the longest in Japan, where the mora consonant was conceived (Kuzobono 2015). Traditionally, the first half of a geminate cluster is not represented as one of the set of obstruents which can become geminate in Japanese, /p t k s ʃ ts ʧ/. Instead, it is common to use an underspecified mora obstruent /Q/ - comparable to Japanese's mora nasal, /N/ - which assimilates the features of the following obstruent (Kuzobono 2015). This tradition is somewhat reflected by the modern representation of geminates as duplicated on the CV- or X-tier, which will be discussed in the fourth chapter.

## 2 Unit Integrity

As shown by the doctrine of the structuralists, the phonemic and/or monosegmental status of geminates is not a nonissue. Indeed, geminates have an ambiguous status, borne out in rules. Muller's examples from a Moroccan Arabic language game are relevant as evidence that geminates are monolithic in that language (Muller 2001). In this language game, consonants in a stem are reversed, and singletons and geminates are treated the same. The data is on the following page.

| **Standard** | **Play** | **Gloss** |
|---|---|---|
| ktb-na | btk-na | ”we wrote” |
| ta-n-ṣum | ta-n-ṃus | ”I am fasting” |
| kubb | bukk | ”he poured” |
| ṃṃ$^{w}$ingina | nnigiṃa | ”he poured” |

Figure 1: Moroccan Arabic

In this data, the geminate and singleton consonants are effectively treated the same - [kubb] generates [bukk], not *[bukb]. The apparent two-consonant portion is treated as one segment.[3] While in this data set geminates are quite clearly singular phonemic units, they also suggests the complexity of their underlying representation. As attested in basically all languages which have them, geminates can cross syllable boundaries (Kuzobono 2017). Syllabification is one element of unit analysis issues, and best addressed now. The syllable-crossing nature of geminates is most similar to tone; both are clearly beyond a feature model. This contrasts with the assumptions of strict linear phonology, where phonemes are discrete. A generalization of the Obligatory Contour Principle suggests that geminates are also dual segments linked to a single quality on a separate representational tier (Curtis 2003).[4] A phonetic realization of, for instance, [fatto] (’fact’) in Italian is underlyingly *fato* with the [t] linked to two consonant slots (as opposed to [fato], ’fate’). What appear to be two identical discrete segments are in fact conjoined on a level above the segments.

[3]One might note that in the third row the geminate quality of the last consonant is retained. This indicates the importance of the analyses below - where weight is comprises a separate tier. One can conjecture that the language game only modifies phonemes, not weight assignments.

[4]This generalization was originally formalized in John McCarthy’s paper in 1979.

# 3 In Rules

The behavior of geminate consonants in rules is the single area of study which has lent the most insight to their phonological status in different languages. Multiple seminal works have focused on the way in which geminates' dual nature is supported by their varying treatment in different languages' analyses. Broadly, geminates are resistant to classical linear rules; not only do they evince their own special phonemic status, they also support autosegmental theory.

## 3.1 Geminate Blockage

A primary issue within this field is geminate blockage. As noted in Kenstowicz and Pyle 1971 it's fairly common for geminates to be the exception to various epenthesis or metathesis rules; this appears to be less an anomaly and more a result of geminates' innate qualities. These observations are followed by William Leben's work (Leben 1980). One of the core data sets discussed by Kenstowicz and Pyle is the following in Sierra Miwok (Kenstowicz and Pyle 1971):

| **Present** | **Past** | **Gloss** |
|---|---|---|
| tuyáːn | tuyánn | "to jump" |
| sakúːk | sakúkk | "to tear" |
| wóʔlu | wóʔull | "to go home" |
| nákpa | nakápp | "to catch up with" |
| hámme | hamé?? | "to bury" |
| ʔúppi | ʔúpi?? | "to dive" |

Figure 2: Sierra Miwok

Kenstowicz and Pyle begin by regarding the geminates as simply consonant clusters. They hypothesize two rules, one of metathesis of a vowel into consonant clusters followed by one of gemination, as well as vowel-shortening before clusters (Kenstowicz and Pyle 1971). This analysis works until one considers the last two rows of data, where instead of an output of *[hamémm] and *[ʔúpipp], as predicted, one finds [hamé??] and [ʔúpi??]. The central issue is that the geminates are not undergoing the metathesis predicted for clusters. While one could conjecture other rules that produce the output, one still has the fundamental problem of the inequality of geminates and clusters. Per the "identity hypothesis," rules that might split up a geminate cluster are less highly valued than a constrained rule (Kenstowicz and Pyle 1971). This phenomenon is crucially explained by the rule that one cannot insert another phoneme into the space on that tier already occupied by one sound attached to two consonant slots (Schein and Steriade 1986); this will be addressed in the next section.

### 3.2 Among Features: Uniform Applicability Condition

Given a sequence of two identical consonants, there are two possibilities for the application of a rule which would apply to one or the other: it is successful, or it fails to apply. In the second instance, geminate status is the agreed-upon cause for this failure (Schein and Steriade 1986; Davis 2011). As in the above section, geminates behave differently from simple sequences, in that they are linked on a tier higher up than that where these feature rules take affect. This is evinced in Schein and Steriade 1986 by data from Tigrinya, which contains an optional rule that "monophthongizes" /ʌj/ sequences when they are in the same syllable. This is shown in the data set on the following page ([j] is transcribed as *y* and [ʌ] as *ä*).

| Standard Form | Alternative Form | Gloss |
|---|---|---|
| sətäy | sətä | ”drink-IMP” |
| sätäy-ka | sätä-ka | ”drink-PERF-2sg” |
| yə-käyd-u | yə-käd-u | ”go-IMPF-3pl” |
| qäyyäd-a | – | ”bind-PERF-3sg” |
| qäyyəd-u | – | ”bind-PERF-GERUND” |
| yə-šäyyəṭ | – | ”sell-IMPF-3sg” |

Figure 3: Tigrinya

Here a [ʌj] sequence in front of another consonant or the word boundary can become [ʌ], but only if the following consonant is not another [j] - the geminate glide is treated differently. A similar process in Tigrinya makes /ʌw/ become /o/, and doesn't apply given an /ww/ sequence. According to Schein and Steriade 1986, this is most sensibly accounted for by an underlying distinction between the dissimilar consonant clusters and the geminate /j/; as in the above section, the geminate is a continuous segment in a way the consonant cluster is not.[5]

As a result of a cross-linguistic pattern of data like this, Schein and Steriade 1986 proposes the ”Uniform Applicability Condition”: given a node *n* and the set of nodes *S* linked to that node, a rule that modifies the contents of *n* must have its structural conditions satisfied by every member of *S*. That is, rules like the one in Tigrinya can't apply to the first half because the second half does not satisfy the structural description. This constraint describes a number of phenomena in other languages with geminates (Schein and Steriade 1986; Davis 2011).

[5]To be fair, /j/ and /w/ aren't fantastic examples of consonants, and indeed in Hayes' model are practically like weightless vowels; that being said, in Tigrinya they pattern like consonants.

# 4 Representations

Modeling of geminate consonants is inextricable from the advances since the 80's and 90's in moraic theory. The original syllable model consisted of an onset and a rhyme, where the rhyme has a subsidiary nucleus and coda. While this representation is amenable to early generative models, it doesn't track with the multitude of rules which seem to operate on different levels of this structure (Curtis 2003). Over time, the hierarchical syllable representation morphed into a more horizontal distribution of weight (Hayes 1989).

## 4.1 Tier Systems

In their respective analyses, authors have chosen several different ways of describing the tier systems that guide the underlying representations of geminates. The main two contenders are a moraic system, where weight is separate from the distribution of phonemic units, and a length system, where weight is folded into segments (Curtis 2003). If one picks the moraic system, one encounters issues with specifications of phoneme place; if one picks the length system, some of the weight assignment is ambiguous. As noted in Stuart Davis' article, even at the level of length diagrams authors must choose between the CV- and X-tier models (Davis 2011). The X-tier models are closer to the moraic ideal, and appear to be more commonly used in contemporary linguistics, but both have a place in the literature.

Famously, Bruce Hayes proposed that a separate "syllabification algorithm" assigns the first half of the geminate to the coda of the first syllable in question, and the second half to the onset of the second - that is, it "flops" onto the following syllable (Davis 2011; Hayes 1989). This model has been criticized by scholars such as Muller, Emily Curtis, and others (Muller 2001; Curtis 2003;

Davis 2011). The separate syllabification rule model fails to account for syllabic consonants with the same weight assignment that behave differently, like in Makhuwa (a Bantu language which contrasts syllabic consonants and geminates), and geminates which seem to have fluid moraic status (Curtis 2003, Muller 2001).

### 4.2 Weight Assignment

Even with the resolution of tier models, one finds several different means of describing weight assignment in the literature. As mentioned above, Hayes in his seminal 1989 article prefers a strict two-slot analysis (Hayes 1989). That is, geminates are always one phoneme linked to two consonant slots on a CV-tier; implicitly, their weight is cross-linguistically constant. Hayes' hypotheses about weight assignment are shown below.

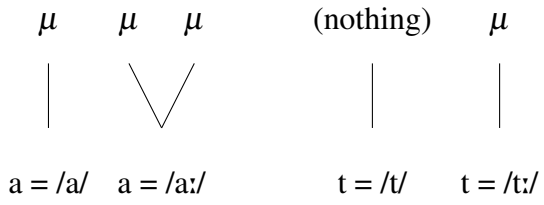


Figure 4: Hayes Mora Model

This analysis, preferable in part for its simplicity, spurred much of the research discussed in the "In Phonological Rules" portion. It has, however, been critiqued. First and foremost, geminates don't appear to have cross-linguistically constant weight. As noted by several authors, in languages with both syllabic consonants and geminates this simple weight model contradicts the predictions of rules (Davis 2011; Curtis 2003). In languages like Chuukese, requirements that a morpheme be bimoraic mean that even word-initial geminates are weighted. Regardless, the weight of geminates

is greater than that of their singleton counterparts. Cuna, Latin, and Iraqi all have some kind of process of degemination which seems to ensure that moras are not too crowded (Curtis 2003). Broadly, the question is how one can judiciously weight geminates while retaining an innate sense of geminates' special qualities. As noted by Muller, word-initial geminates like those discussed in 1.1 (in languages besides Chuukese) are not definitively moraic (Muller 2001). However, it is possible to adapt the moraic model to these forms, if one considers a multi-tiered model of weight (Davis 2011). Generally, the weight model of geminates is more prominent, and within that model the geminates-as-moraic-consonants description is largely accepted.

## 5 Conclusion

In sum, geminate consonants can be represented in a variety of ways. All of the current systems for describing geminates' weight are autosegmental. Even so, a featural representation was popular in the early days of generative phonology, and in traditional Japanese phonology, geminate obstruents were represented as a sequence of an underspecified segment /Q/ and the base consonant. As it stands, a general interpretation of Bruce Hayes' model of the geminate consonant's attachment to a moraic unit seems to hold up the best; geminates are consistently differentiated from their singleton counterparts. They are also distinct from consonant clusters, even though methods of both transcription and modeling in linguistics can be somewhat confusing on this point. Geminates relate to the interface between phonetics and phonology, as well as underlying structures and surface realizations; hence, their representation has implications for many different manifestations of the autosegmental model.

## References


Curtis, Emily (2003). *Geminate weight: Case studies and formal models*. eng. URL: http://search.proquest.com/docview/305305329/.

Davis, Stuart (2011). "Geminates". In: *The Blackwell Companion to Phonology*. Wiley-Blackwell. Chap. 37, pp. 1–25. ISBN: 9781444335262. DOI: 10.1002/9781444335262.wbctp0037.

Hayes, Bruce (1989). "Compensatory Lengthening in Moraic Phonology". In: *Linguistic Inquiry* 20, pp. 253–306. URL: https://linguistics.ucla.edu/people/hayes/papers/HayesCompensatoryLengthening1989.pdf.

Kenstowicz, Michael and Charles Pyle (1971). *On the phonological integrity of geminate clusters*. eng.

Kuzobono, Haruo, ed. (2015). *Handbook of Japanese phonetics and phonology*. eng. Handbooks of Japanese language and linguistics, volume 2. Boston: De Gruyter Mouton. ISBN: 1-5015-0059-7.

— ed. (2017). *The phonetics and phonology of geminate consonants*. eng. First. Oxford studies in phonology and phonetics; 2. Oxford: Oxford University Press. ISBN: 9780198754930.

Leben, William R. (1980). "A Metrical Analysis of Length". In: *Linguistic Inquiry* 11.3, pp. 497–509. ISSN: 00243892, 15309150. URL: http://www.jstor.org/stable/4178177.

Maddieson, Ian and Kristin Precoda, eds. (1989). *UPSID*. Cambridge University Press. URL: http://web.phonetik.uni-frankfurt.de/S/S0097.html.

Muller, Jennifer (2001). "The phonology and phonetics of word-initial geminates". eng. PhD thesis. Ohio State University. ISBN: 0493440178. URL: http://search.proquest.com/docview/251637564/.

Schein, Barry and Donca Steriade (1986). “On Geminates”. In: *Linguistic Inquiry* 17.4, pp. 691–744. URL: http://www.jstor.org/stable/4178513.

Swadesh, Morris (1937). “The Phonemic Interpretation of Long Consonants”. In: *Language* 13.1, pp. 1–10. URL: http://www.jstor.org/stable/409167.