8. Phonological Quantity and Multiple Association

DAVID PERLMUTTER

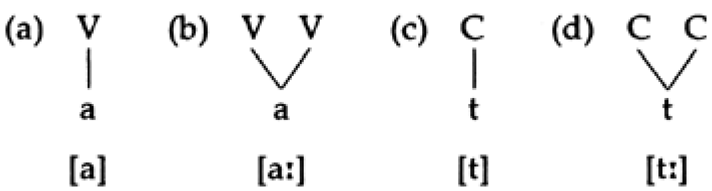
Subject	Theoretical Linguistics » Pholonogy
DOI:	10.1111/b.9780631201267.1996.00010.x

0 Phonological Quantity: From a Feature to Multiple Association

Languages that have contrasts between short and long vowels and, in some cases, between single and geminate consonants, are said to have contrasts in phonological quantity. Luganda, for example, exhibits both types of contrast, e.g., *wela* "refuse" versus *weela* "rest" and *yiga* "learn" versus *yigga* "hunt" (Tucker 1962), where length is transcribed by double letters. The framework of Chomsky and Halle (1968) had relatively little to say about quantity, assuming it could be handled by means of a feature such as [+Long].

In his work on Arabic verb morphology, McCarthy (1979a, 1981a) took an important step, separating not only consonant sequences (verb stems) from vowel sequences as different morphemes on separate tiers, but also the pattern of vocalic and consonantal positions as a separate morpheme on a separate tier. Since Arabic uses quantity to distinguish morphologically distinct forms of a verb (e.g., *katab* vs. *kaatab* vs. *kattab*), McCarthy represented contrasts in quantity as contrasts in the number of V or C slots with which a melodic unit is associated:

(1)



This analysis autosegmentalizes segmenthood, separating it from melodic features onto a "skeletal tier" or "skeleton," much as autosegmentalization of tone separated tonal features from melodic features of individual segments. Since linking of a single melodic unit to more than one C or V slot ("multiple association" or "multiple linking") is realized as length, the skeleton is sometimes called the "timing tier."

Arguments for this autosegmental analysis of quantity must eliminate two alternatives: a feature analysis of quantity and the representation of quantity as a sequence of two identical melodic units. For example, in arguing for multiple association of long vowels in Finnish, Prince (1984) shows an environment in which single vowels are deleted and long vowels are shortened. Multiple association yields the correct result if a skeletal V slot is deleted (with concomitant deletion of associated melodic material not linked to any other slot). Having two slots for long vowels is crucial to capturing the generalization. Analogously, a rule simplifying clusters deletes a consonant before a sequence of two Cs and before geminates. Multiple association of geminates allows the environment to be stated as preceding a sequence of two C slots. These phenomena argue against an analysis of quantity as a single melodic unit with the feature [+Long], but they could be handled if long vowels and geminates were represented as sequences. That alternative is eliminated by restrictions on consonant clusters. Geminates are systematically allowed to end clusters where other sequences of two obstruents are not. To distinguish sonorants from obstruents, these constraints must refer to features of melodic units. Multiple association provides the right representation to capture the fact that geminates count as single consonants with respect to these cluster constraints. Prince also argues that geminates do not behave like clusters with respect to consonant gradation, which crucially refers to melodic features. The generalization is clear: phenomena that count the number of segments "see" long vowels and geminates as sequences of two segments, while phenomena that look at their feature content "see" only one. This is precisely what the representations in (1) provide.

Another type of argument for multiple association appeals to the fact that devices familiar from the autosegmental analysis of tone can account for phonological quantity once segmental slots are placed on a separate tier from melodic features. One such device is autosegmental spreading, used to account for compensatory lengthening in a CV framework (Ingria 1980; Wetzels and Sezer 1986). For example, in Luganda (Clements 1986a), concatenation of the plural prefix *ba-* with the stem *ezi* "sweeper" yields *beezi*. The problem is how to make lengthening of the stem vowel a consequence of deletion of the prefix vowel. Multiple association does this:

(2)

Deletion of the *a* leaves an empty V slot that associates with the neighboring vowel by autosegmental spreading, familiar from the spreading of tone. Since multiple association is realized as length, this analysis correctly makes lengthening of the *e* depend on deletion of the preceding vowel.

Evidence for multiple association naturally gave rise to the question of why representations should have this form. Here, too, appeal was made to a device used for tone: the Obligatory Contour Principle or OCP (see chap. 12, this volume), which rules out identical adjacent autosegments on the same tier. Since a sequence of two identical melodic units violates the OCP, long vowels and geminates must be represented with only one. Multiple association is then the only available representation that accounts for the ways they behave like sequences of segments. The evidence that rules referring to feature content "see" only one melodic unit in these cases (as in Finnish) supports this use of the OCP.

Arguments for the representation of quantity as multiple association to segmental slots figured prominently among the arguments for a CV skeleton, which included both one-many and many-one skeleton-melody associations, morphological templates like those of McCarthy (1979a, 1981a), and "empty" skeletal slots unassociated with melodic content (Clements and Keyser 1983).

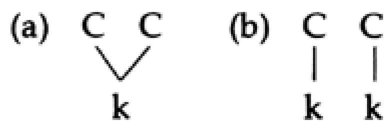
1 The Special Behavior of Geminates: Integrity and Inalterability

With multiple association, attention again focused on why certain rules do not apply to geminates in the ways the representation would lead one to expect (Kenstowicz and Pyle 1973; Guerssel 1977, 1978). Geminates are said to exhibit *integrity* where epenthesis rules fail to split them up, and *inalterability* where rules fail to apply to them. Since multiple association represents geminates with only one melodic unit, the representation explains why a feature-changing rule cannot change only "half" of a geminate: application of the rule would affect the features associated with both segmental slots, resulting in a change in both "halves" of the geminate. What remains to be explained is why certain rules fail to apply to geminates at all. These are cases of inalterability.

Recent years have seen several attempts to explain the behavior of geminates. The pioneering studies of Hayes (1986b) and Schein and Steriade (1986) focus on the double association lines in geminates' representations, requiring that they both be explicitly mentioned in the rule in order for the rule to apply.[1] This approach has been criticized by Scobbie (1993) and Inkelas and Cho (1993). Scobbie attributes inalterability to the ill-formedness of a rule's output rather than to the way the rule matches against input. Inkelas and Cho argue that rules respecting inalterability are structure-filling rather than structure-changing, and that inalterability effects are also found with nongeminates. Both studies argue against constraints aimed specifically at the properties of representations with multiple association.

The special behavior of geminates has been harnessed to highlight the power of representations in achieving explanation. A key distinction is that between "true geminates," represented with multiple association, as in (3a), and "apparent geminates" (or "fake geminates"), which, though phonetically indistinguishable from them, are represented as two separate melodic units linked to separate C slots in the skeleton, as in (3b):

(3)



If the OCP is a condition on lexical representations, there can be no morphemeinternal apparent geminates. Apparent geminates are necessarily heteromorphemic, their existence indicating that (at least in some cases) the OCP does not rule out identical adjacent melodic units across morpheme boundaries.[2] Further, wherever geminates arise through autosegmental spreading (as in compensatory lengthening and total assimilation), a doubly linked geminate representation (a true geminate) results. These assumptions, then, assign multiply linked representations to tautomorphemic geminates and those that arise through spreading. Such true geminates contrast dramatically with apparent geminates in Tigrinya (Schein 1981; Kenstowicz 1982; Schein and Steriade 1986).

The key Tigrinya rule spirantizes stops that are preceded by a vowel, even across morpheme boundaries (e.g., *gäzaxa* < *gäza* + *ka* "your house"). True geminates fail to undergo spirantization, regardless of whether they are tautomorphemic (*fäkkärä* "boast (3d sg. masc. perf.)", dervied by total assimilation (*yəqqayädu* < *y* + *t* + *qayädu* "attach [imperf.]"), or by spreading of stem melodic material to an empty C slot in a suffix (*yəbarəkko* <*yəbarək* + Co "bless [3d sg. masc. jussive]"). The suffix in the last example is analyzed with an empty C slot because it characteristically causes gemination of a stem-final consonant, accomplished through spreading. These true geminates all exhibit inalterability in failing to spirantize. They contrast with the apparent geminates resulting from suffixation of -*ka* (with a specified *k*), which undergo spirantization (e.g., *baräxka* < *baräk* + *ka* "you blessed"). Given inalterability of multiply associated geminates, the contrast between the three types of true geminates (which do not spirantize) and the apparent geminates (which do) strongly supports the elements of the theory that dictate multiple association for the former.

2 The Prosodic Theory of Phonological Quantity

The theory discussed so far represents phonological quantity as the association of more than one C or V slot in the skeleton with a single melodic unit. Because it makes reference to segmental slots but not to higher prosodic structure, I call this the segmental theory of quantity.[3] Recent research has led to the elaboration of two main types of prosodic theories of quantity. One associates melodic units with segmental slots, replacing C and V slots with generalized X slots and relying on a richer theory of syllable structure that includes a nucleus node, generally representing "heavy" syllables (cf. below) as having a branching rhyme or nucleus node (Hayes 1980; Prince 1983; S. Anderson 1984; Levin 1985a; Lowenstamm and Kaye 1986; Kenstowicz and Rubach 1987; and others). The other associates melodic units with mora nodes in prosodic representations (Hyman 1985; Hayes 1989a; McCarthy and Prince 1986, 1990a, 1993; and others). Space limitations prevent discussing the former or differences between them here.[4] The important point of agreement is that an understanding of quantity requires consideration of prosodic structure above the segmental level.

The CV theory of quantity leaves some important questions unanswered. What is the relative quantity of long vowels and geminate consonants? Are other strings of segments quantitatively equivalent to either of these? Is there a common unit of measure of quantity in terms of which relative quantity can be stated? Some light is shed on these questions by the distinction many languages make between "light" and "heavy" syllables" (Newman 1972; Hyman 1985; Hayes 1994), perspicuous in "quantity-sensitive" stress systems, where heavy syllables tend to attract stress.[5] Heavy syllables can be of two types: (1)syllables with a long vowel or diphthong, and (2) closed syllables, or syllables closed with certain consonants. Language-particular restrictions determine which syllables of these types are heavy. Syllables closed with a geminate fall under (2). This raises the question of what (1) and (2) have in common that makes both heavy, and how this is to be represented formally - a question on which recent studies have shed light.

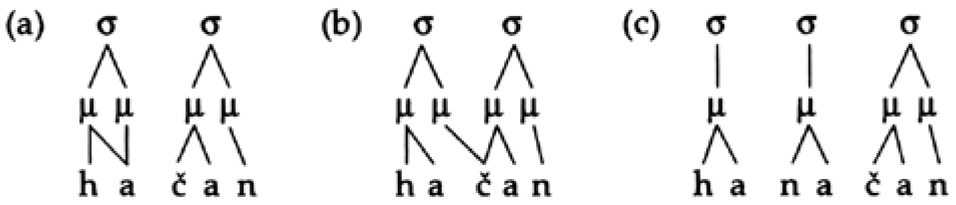
Poser (1990) shows that the class of stems to which -*čan* can be suffixed in Japanese hypocoristics cannot be characterized in terms of the number of segmental slots or melodic units. From the name *hanako*, for example, three hypocoristics can be formed:

(4)

(a) haa-čan (Suffix -*čan* is preceded by 2 melodic units, 3 segmental slots)
(b) hač-čan (Suffix -*čan* is preceded by 2 melodic units, 3 segmental slots)
(c) hana-čan (Suffix -*čan* is preceded by 4 melodic units, 4 segmental slots)
(d) *-čan

For satisfying hypocoristic requirements, a syllable with a long vowel is equivalent to one closed with a geminate consonant, and both are equivalent to a sequence of two light (open, short-voweled) syllables. There is a common unit of measure - the mora: a light syllable has one, and a syllable with a long vowel or closed with a geminate has two. The forms in (4) are thus:

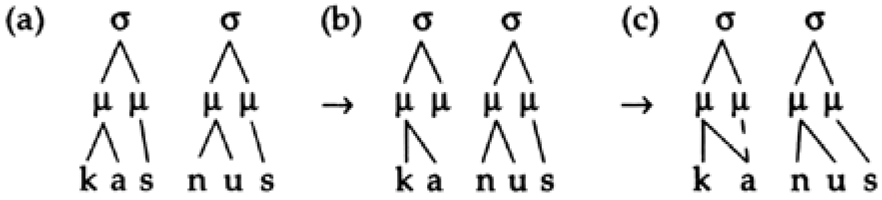
(5)



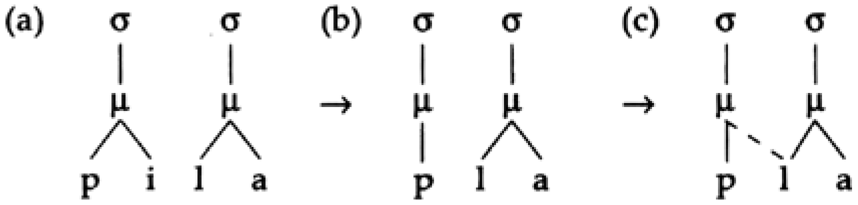
A vowel linked to two moras is long; a consonant that is moraic in one syllable and also the onset of the next is realized as a geminate. A hypocoristic stem in Japanese must consist of two moras. A long vowel or a geminate consonant contributes a second mora to the syllable, which is consequently heavy. Japanese hypocoristics bear on quantitative equivalence: with the mora as unit of measure, quantity in vowels and consonants is equivalent, for each adds a mora to the syllable, making it bimoraic. Further, a syllable with either vocalic or consonantal quantity is quantitatively equivalent to a sequence of two light syllables.

Hock (1986) and Hayes (1989a) use compensatory lengthening to argue for moraic representations. As has long been recognized, lengthening of a vowel can result not only from deletion of another vowel, but also from deletion of a vowel. Hayes exploits the autosegmental analysis of quantity, but with multiple association to moras rather than to segmental slots. Deletion of a vowel or consonant leaves an empty mora node, whose associaton with another melodic unit results in quantity.[6]

(6)



(7)



In (6), from Latin, where coda consonants are moraic and closed syllables are consequently heavy (bimoraic), deletion of the coda *s* results in association of *a* with the stranded mora, yielding *kainus*. In (7), attested in several African languages and as a fast speech rule in French, deletion of *i* results in linking of the following *l* with the stranded mora, yielding *plla*, in which the geminate is nucleus of the first syllable. Such cases are another source of evidence that what makes a vowel long is also what makes a consonant a geminate: association to an additional mora.

These cases do not require moraic representations, however; they can be handled by the theory with generalized X slots. Hayes' principal argument that compensatory lengthening is association to a stranded mora is based on the asymmetry between moraic and nonmoraic consonants. Deletion of an onset consonant does not cause vowel lengthening because onset consonants are not moraic. Deletion of a coda consonant can cause vowel lengthening, but only in languages where closed syllables are heavy, indicating that the coda consonant is dominated by a mora. The analysis further predicts that in languages where only certain coda consonants are moraic, only deletion of *moraic* coda consonants can cause compensatory lengthening. The general claim is that moraic structure, like tonal autosegments in tone languages, is stable under deletion of melodic material, which leaves the mora count intact. Association of the mora stranded by deletion with another melodic unit results in phonological quantity on that segment: compensatory lengthening. Phonological quantity thus cannot be understood in terms of segmental slots alone. Moraic structure, left intact under segmental deletion, shows up as quantity on individual segments with which an additional mora is linked.

Compensatory lengthening and the examples in (4) point in the same direction. First, they bring out the equivalence of different manifestations of phonological quantity and show that it can be captured in terms of the mora. Second, both are cases of *prosodically motivated quantity*: quantity appears in (4) and (5) to satisfy a bimoraic prosodic template, and in (6) and (7) because preexisting moraic structure is preserved and reassociated. Such phenomena make clear that phonological quantity is an aspect of prosodic (rather than segmental) structure.[7]

Co-occurrence restrictions between long vowels and geminate consonants show that quantity is intimately linked to syllable weight. Japanese is typical in that a long vowel can be preceded by a geminate consonant (e.g., *gakkoo* “school”) but not followed by one (**gaakko* is impossible). A theory that looks only at segmental slots and not at higher prosodic structure does not see the key to an explanation: these words are syllabified as *gak.koo* and **gaak.ko*. It fails to capture the generalization that a long vowel cannot be the nucleus of a closed syllable. The syllable is thus the domain in which the co-occurrence restrictions can be explained. The explanation rests on the fact that Japanese exhibits a binary opposition between light (monomoraic) and heavy (bimoraic) syllables. Since a long vowel is bimoraic and a coda consonant contributes another mora to the syllable, a closed syllable with a long vowel (such as the first syllable of **gaak.ko*) would be trimoraic. By positing a bimoraic upper bound on syllables in Japanes morphemes,[8] the prosodic theory explains why a long vowel cannot be followed by a geminate consonant. It also explains two other facts: (1) why a long vowel cannot occur in other closed syllables, e.g., alongside *hon* "book", **hoon* is impossible;[9] and (2) why Japanese does not manifest three-way length contrasts, e.g., alongside *kite* "come!", *kiite* "listen!", and *kitte* "stamp", **kiiite* and **kittte* are impossible.[10] These aspects of the distribution of quantity are explained under the hypothesis that the syllable weight opposition is binary.

A prosodic theory of quantity that links melodic units to moras rather than segmental slots automatically explains geminate integrity. Under the segmental theory, it is necessary to prevent epenthesis rules from inserting another segmental slot between the two segmental slots used to represent quantity. With moras rather than segmental slots in the skeleton, a geminate consonant is linked to a mora in the previous syllable and serves as onset of the next, as in (5b). A geminate is not represented with two segmental slots, so there is no way anything could be inserted in its "middle." By the same token, as under the segmental theory, there is no way a rule could alter only "half" of a doubly linked geminate. However, it is still necessary to explain the cases where phonological rules fail to apply to geminates at all. Essentially the same range of explanations is available as under the segmental theory, mutatis mutandis.

An apparent anomaly of geminates' distribution lies in the fact that in many languages, certain consonants cannot close a syllable, but if they are geminates they can. In Japanese, for example, a syllable cannot be closed with an obstruent, but a geminate obstruent (e.g., *kitte*) is allowed. The same holds in Hausa. Recent proposals attribute geminates' ability to close a syllable under these conditions to the fact that they are both coda of one syllable and onset of the next. Itô (1989) proposes a filter that rules out codas with Place features, to which Hayes' (1986) Linking Constraint makes geminates immune because they are doubly linked. Under Yip's (1989) proposal, geminate codas are immune to coda constraints that ban Place features because they have no Place features of their own, "borrowing" them from the following onset. Goldsmith (1989, 1990) proposes an explanation in terms of autosegmental licensing, the idea that all features of a segment must be licensed by either the syllable node or the coda node under which they occur. The restrictions on coda consonants are interpreted in terms of the coda's restricted ability to license features. To explain geminate obstruents' ability to close a syllable, Goldsmith appeals to the fact that they are both coda of one syllable and onset of the next syllable, hence it need not be licensed as coda; its (unlicensed) appearance as coda is parasitic on its licensed appearance as onset.

Using the framework of harmonic phonology (Goldsmith 1993a), Goldsmith (1989, 1990) further exploits autosegmental licensing to explain geminates' integrity and inalterability. The key harmonic assumption is that phonological rules apply only in order to increase representations' wellformedness with respect to a set of word-level phonotactics. On this view, rules that weaken coda consonants do so in order to eliminate features that cannot be licensed by the coda, so that wellformed representation results. A representation with a geminate obstruent closing a syllable is well-formed because the obstruent is licensed as onset of the next syllable. Since rules apply only to increase representations' well-formedness with respect to a phonotactic, they need not apply to geminates. Hence they do not. In this way, Goldsmith's proposal links geminates' integrity and inalterability to their ability to appear in positions where the corresponding nongeminate consonants cannot.

3 Conclusions: Phonological Quantity, Syllable Weight, and the Mora

The past fifteen years' research has significantly advanced the understanding of phonological quantity. Quantity has autosegmental properties reminiscent of the stability and spreading of tone in tone languages. This is captured by autosegmental representations that put quantity on a separate tier from melodic material. Such representations also account for the ways segments with quantity behave like a single melodic unit and for the contrast between true and apparent geminates in languages such as Tigrinya.

What does the tier with which melodic units are associated consist of? The study of quantity provides evidence that melodic units are associated with moras in prosodic representations: the unit of measure of quantity is the mora, in terms of which it is possible to state quantitiative equivalances and differences between vowels and consonants, and between "long segments" and other strings of segments. Association with moras also accounts for the stability and spreading of quantity, for moras are stable and spread under deletion of melodic material; association of melodic material to a stranded mora results in quantity (compensatory lengthening is indirect, mediated by moraic structure. Deletion causes compensatory lengthening because it results in a stranded mora. Association of a segment with an additional mora results in quantity.

What, then, is the mora? The glossary in Bright (1992) defines it as a unit of length, and the recent phonological literature generally calls it a unit of both length and weight. The mora, however, is not a unit of length in the simple sense that each mora in a phonological representation represents an equal timing unit. If that were so, a moraic nongeminate consonant and a short vowel would have the same duration, and a long vowel would have exactly twice the duration of these. Nongeminate coda consonants would also be significantly longer in languages where closed syllables are heavy than in those where they are not. Sequences of nongeminate coda consonants in such languages would be no longer than a single coda consonant, since the mora count would be the same. In this sense it is misleading to think of moras as constituting a "timing tier."

Quantity is predictable, however, from representations in which melodic units are linked to moras:

(8)

(a) A vowel dominated by two moras is realized as a long vowel.
(b) A consonant dominated by two moras is realized as a geminate.

With representations like those in (5), the right results are obtained. (8) is essentially a cross-linguistic summation of language-particular phonetic rules that specify the duration of long vowels and geminate consonants in individual languages. Syllables with a long vowel and those closed with a geminate can be quantitatively equivalent without having the same duration. The mora is not a unit of length, but a unit of measure from which quantity is predictable.

What, then, does the mora measure? There is evidence that the mora is a unit of weight, and that the quantitative equivalences discussed above are consequently equivalences of syllable weight. The key fact bearing on this question concerns the distribution of segments with quantity in phonological strings. We have seen that a long vowel cannot be followed by a geminate consonant in languages such as Japanese, and that this can be explained in terms of the binarity of the syllable weight opposition in those languages, which imposes a bimoraic upper limit on the syllable. Syllable weight is thus the concept that explains co-occurrence restrictions among segments with quantity. The maximum number of moras per syllable is that needed to express a language's contrasts in syllable weight, although a larger number of segments are capable, a priori, of expressing quantity. This is evidence that the mora is a unit of weight.

Another piece of evidence that the mora is a unit of weight comes from the cross-linguistic generalization that phonological quantity is found only in languages that distinguish between light and heavy syllables (Hayes 1989a). If the mora is a unit of weight, only such languages can have structures that (8) will interpret as long vowels or as geminate consonants. The recent literature has thus been correct in characterizing the mora as a unit of weight, but incorrect in characterizing it as a unit of length. Weight is interpreted as quantity under the conditions specified in (8). "Phonological quantity" turns out to be a cover term for oppositions whose locus lies in contrasting associations of melodic material with units of weight (moras) in prosodic structure.

*I am indebted to Juliette Blevins, Philip Le Sourd, and Moira Yip for comments on a draft of this chapter, and to Matthew Gordon for helpful dicussion. Errors and inadequacies are mine alone.

1 This is also the approach of Goldsmith's (1990, p. 39) Conjunctivity Condition. One difference between Hayes's condition and the others is that the latter focus only on the assocation lines in the *target* of a rule, allowing a multiply associated autosegment to be the trigger.

2 Only if they are on the same tier would the domain of the OCP be at issue. The OCP has given rise to a number of issues that cannot be broached here. Among them: (1) Is the OCP a universal principle, a cross-

linguistic tendency, or a cover term for a number of language-particular rules? (2) Is its domain lexical and hence morpheme-internal, or is it broader? (3) Does it merely rule out certain representations as ill-formed, or does it "actively" convert identical adjacent same-tier autosegments to multiply linked autosegments? (2) interacts crucially with the Morphemic Tier Hypothesis (McCarthy 1979a, 1981a, 1982b, 1986b), which in crucial cases places distinct morphemes on separate tiers, where they are immune to potential cross-morphemes OCP effects. See McCarthy (1986b), Odden (1988), Yip (1988a), Goldsmith (1990, pp. 309-318), and Odden, chapter 12, this volume, for discussion.

3 This is really an abstraction for expository purposes since the CV skeleton has always been integrated into a theory of the syllable (Clements and Keyser (1983).

4 See Broselow, chapter 5 and Blevins, chapter 6, this volume, and the references cited there. Space limitations also prevent discussion here of the use of multiple association for the representation of biliteral roots in Semitic languages with so-called "long-distance geminates" (McCarthy 1981a, 1982b, 1986b) (see Hoberman, chapter 30, this volume).

5 See Kager, chapter 10, this volume, and the references cited there.

6 Onset consonants have been placed under the vowel nucleus's mora node here to conform with the Strict Layer Hypothesis of Selkirk (1984b).

7 The prosodic theory must be able to represent underlying contrasts in quantity. One proposal is that of Hayes (1989a), who represents short vowels with one mora, long vowels with two, and geminate consonants with an underlying mora.

8 Poser (1984b) shows that the handful of exceptions arise only across only morpheme boundaries.

9 A handful of loan words (e.g., *toronboon* "trombone") are the only exceptions.

10 It is tempting to posit that the bimoraic upper limit is universal, and to make it follow from a putative principle of prosodic theory: the branching of prosodic constituents is *maximally binary* (see Kager, chapter 10, and McCarthy and Prince, chapter 9, this volume, and references cited there). Since syllable nodes dominate mora nodes in prosodic structure and no other node type can intervene (cf. the Strict Layer Hypothesis of Selkirk 1984b), there can be at most two moras per syllable. However, putative cases of trimoraic syllables militate against such an explanation (cf. Hayes (1989a, pp. 293-297) and the references cited there).

Cite this article

PERLMUTTER, DAVID. "Phonological Quantity and Multiple Association." *The Handbook of Phonological Theory*. Glodsmith, John A. Blackwell Publishing, 1996. Blackwell Reference Online. 31 December 2007 <http://www.blackwellreference.com/subscriber/tocnode?id=g9780631201267_chunk_g978063120126710>

Bibliographic Details

The Handbook of Phonological Theory

Edited by: John A. Glodsmith
eISBN: 9780631201267
Print publication date: 1996