# Phonetics-phonology interplay in moraic theory: Evidence from Finnish dialectal gemination ${ }^{1}$ 

## 1 Introduction

### 1.1 Finnish: a little background

Finnish is a "full-fledged quantity language" (Suomi et al 2008:39). Vowels are contrastive for length in all positions, and consonants are contrastive for length intervocalically, regardless of primary stress, which is always initial.

The examples in (1) from Standard Spoken Finnish show that minimal sets for length abound, both lexical and derived: ${ }^{2}$

```
a. muta 'mud'
b. mutaa 'mud' (partitive case)
c. mutta 'but'
d. muuta 'other' (partitive case)
e. muutta 'other' (abessive case)
f. muuttaa 'to change'
```

These are the facts for Standard Spoken Finnish (SSF), but matters can be complicated by other processes in certain varieties of Finnish. With that in mind, the goals of this talk are as follows:

- to give an overview of two dialectal processes, "primary gemination" and "second mora lengthening"
- to show a relationship between the two, where a phonetic process (second mora lengthening) feeds phonological gemination
- to present a formal model of this interaction within moraic theory (Hayes 1989)
- to discuss the implications for the study of quantity and the phonetics-phonology interface

[^0]
### 1.2 Primary Gemination

A number of dialects of Finnish have a process known as "primary gemination" (yleisgeminaatio), whereby a singleton consonant is geminated in between a stressed short vowel and an unstressed long vowel. This is schematized in (2).

## (2) Primary gemination

$$
\mathrm{CV́}_{\alpha} \mathrm{VV} \rightarrow \mathrm{CV́}_{\alpha} \mathrm{C}_{\alpha} \mathrm{VV}
$$

The result of this is that there exist alternations in these dialects which do not exist in "Standard" varieties of Finnish. Consider the nominative and partitive case forms of the word sana 'word' in a "Standard" variety (one without the gemination process) and a Geminating variety:

|  | "Standard" | Geminating |  |
| :--- | :--- | :---: | :--- |
| a. /sana/ | sana | sana | 'word (nominative case)' |
| b. | /sana-a/ | sanaa | sannaa | 'word (partitive case)'

The alternation (underlined in 3) presents itself when the second vowel in a CVCV root becomes long as the result of suffixation. Thus the root appears as geminated in the partitive case in a Geminating variety.

### 1.3 Some clarifications about applicability

It should be noted that primary gemination is quite widely applicable in varieties that have it, even though my examples here are mostly with the word sana 'word', however, the process is fully productive. (4) shows some additional examples: ${ }^{3}$

```
a. vika 'fault' vikkaa 'fault (partitive case)'
    b. pata 'pot' pattaa 'pot (partitive case)'
    c. raha 'money' rahhaa 'money (partitive case)'
    d. data 'data' dattaa 'data (partitive case)'
    e. mopo 'moped' moppoo 'moped (partitive case)'(cf. non-dialectal mopo-a)
```

In (4c), we find that $/ \mathrm{h} / \mathrm{can}$ be geminated, despite the fact that the language contains a lexical geminate in only one word, hihhuli 'religious fanatic' (Suomi et al 2008:41).

In (4d), we see that it can apply to even a relatively new word such as data, even in more technical usage (for example in computer programming).

Finally, in (4e), we see that it can apply to a long vowel which is derived from the coalescence of two separate vowels: the squence /oa/ becomes a long /oo/, which results in the creation of the environment for gemination.

[^1]
### 1.4 Second Mora Lengthening (SML)

Suomi and Ylitalo (2004) posit a regular process of phonetic lengthening in a number of varieties of Finnish, whereby the second mora of a word is lengthened, provided it is voiced. They call this "Second Mora Lengthening". I term the process SML, which I state in (5):

## (5) Second Mora Lengthening (SML)

Attribute additional weight and length to the second mora of a phonological word
One way in which SML manifests itself is in a phonetic feature known as the half-long vowel (puolipitkä vokaali).

With the half-long vowel, the second (unstressed) short vowel in a CVCV word is actually phonetically longer than the first (stressed) short vowel. I indicate this in (6) with a length mark:

$$
\begin{array}{ccc}
\text { "Standard" variety } & \text { Half-long vowel }  \tag{6}\\
\text { sana } & \text { sana: } & \text { 'word (nominative case)' }
\end{array}
$$

The actual phonetic duration of the half-long vowel (and thus the amount of additional length attributed by SML) varies slightly from dialect to dialect, but it is generally about 1.5 times the duration of the stressed short vowel in the first syllable, but nevertheless shorter than a full long vowel, which is around twice the duration of a short vowel in the same position.

This is illustrated in (7), which shows the segmental durations of CVCV words in three varieties of Finnish. The "Oulu" and "Turku" varieties have the half-long vowel (and thus SML), while the "Tampere" variety does not (data from Ylitalo 2009:52).
(7) Average Segmental Duration in Three Varieties (milliseconds)

|  | $\mathrm{C}_{1}$ | $\mathrm{~V}_{1}$ | $\mathrm{C}_{2}$ | $\mathbf{V}_{2}$ | Total Duration |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Oulu | 65 | $\mathbf{5 7}$ | 71 | $\mathbf{8 7}$ | 280 |
| Turku | 71 | $\mathbf{5 6}$ | 85 | $\mathbf{6 5}$ | 277 |
|  |  |  |  |  |  |
| Tampere (no SML) | 73 | $\mathbf{5 8}$ | 72 | $\mathbf{5 3}$ | 256 |

The data in (7) are a representative sample, but numerous other phonetic studies show similar patterns for varieties with the half-long vowel, including Suomi and Ylitalo (2004), Ylitalo (2004), Lehtonen (1970), and Wiik and Lehiste (1968).

## 2 Gemination and SML: Evidence for the correlation

For the analysis I present below, I posit that the phonetic process SML feeds phonological alternation of primary gemination. In order for this to work, there must a correlation between varieties with SML and those with primary gemination.

One source of data on which areas have what is Kettunen (1940), a seminal dialect atlas of Finnish. Kettunen's Map 2 shows the areas with gemination processes (including primary gemination), while Map 198 shows areas with the so-called "half-long vowel" (puolipitkä vokaali). But the maps do not line up! Gemination processes abound in many regions of Finland, while the half-long vowel is only present in a small area of the south-west part of the country.

How can we reconcile this?

Kettunen (1940) was a study conducted by Lauri Kettunen, by traditional field work (while driving around on a motor cycle, according to the accompanying volume's preface). Because there was no field equipment being used, no phonetic measurements were being taken. As we have already learned, the half-long vowel is a phonetic process, which cannot be switched off even by a speaker who has it. Is this something a field-worker's ear can even detect?

Wiik and Lehiste (1968:572) shed some light on this:
"The half-long vowel is easily heard in the Standard Finnish speech of those from the South-West, but not in the speech of, for example, those from Savo. This may be due to the fact that the term "half-long vowel" as generally used in Finland refers not only to the duration of the vowels, but also to other characteristics, such as fundamental frequency and intensity contours of the words in which the half-long vowel appears. It is possible that these other characteristics differ in the two dialect areas in which the half-long vowel was found to occur."

So the half-long vowel of the south-west - exactly the one which Kettunen was reporting on-has more going on with it prosodically, and is clearly salient enough to report based on "old-fashioned" field work. Wiik and Lehiste's (1968) study, on the other hand, involved phonetic measurement of half-long vowel duration, found it to be far more common in Finnish dialects.

The map in Figure 1 shows a composite of Kettunen's (1940) areas with gemination processes and the locations of the home towns of the speakers in Wiik and Lehiste's (1968) study. Here the correlation between the two processes (gemination and SML) is much stronger!

Figure 1: Combination of Kettunen (1940) map 2 and Wiik and Lehiste's map of speaker home towns (1968).


- Circles indicate home towns of speakers with the half-long vowel.
- Stars indicate home towns of speakers without the half-long vowel.
- The question mark indicates the home town of a speaker for whom the presence/lack of the half-long vowel was an "intermediate case".
- Shaded areas lack gemination, according to Kettunen (1940).


## 3 Analysis

### 3.1 A previous account

The analysis I present below relies on a connection between primary gemination and the half-long vowel (and therefore SML), and such a connection has in fact been posited before. Nahkola (1987:25) notes that there seems to be a connection between the two. The explanation offered by Nahkola, however, is historical in nature.

The long vowel in the second syllable of words like sanaa 'word (partitive case)' derives historically of elision of a intervocalic /ð/:

```
sana-a < *sanaða
```

By this explanation, the application of a half-long vowel to the second short vowel in the historical *sanað a resulted in a CV́CVV word with an overlong (ylipitkä) vowel after elision of /ð/ took place.

The presence of an overlong vowel in the unstressed syllable created some sort of weight inbalance, the story goes, which was corrected by geminating the intervocalic consonant. I give an historical "derivation" of this explanation in (9):
(9) Deriving primary gemination historically
*sánaða 'word (partitive case)'
a. Half-Long Vowel sána:ða
b. Elision sána:a
c. Primary Gemination sánnaa (attested modern form)

There are several shortcomings to this explanation. First, it is not spelled out enough in terms of what exactly the imbalance is which caused gemination. We normally wouldn't expect compensatory lengthening for the deletion of an onset as we have here (see Topintzi 2010 for discussion of such controversial cases).

Second, Nahkola notes that this explanation only works provided that the half-long vowel occurred before elision of the intervocalic / $\delta /$, which may or may not have been the case.

The bigger problem for such an analysis is the way it views historical change. That is, gemination isn't viewed as a phonologically conditioned alternation, but rather as a change which either did or did not apply to individual forms. In other words, it does not consider the notion that morphologically complex words are built by combining different morphemes, but rather that words simply have different forms, which are free to change diachronically.

Most importantly, how do we account for the productivity of primary gemination, where it applies to loanwords newer than elision? What about underlying CVCVV sequences, to which it also applies? We need a generative model!

### 3.2 Unifying the processes

I present a model which avoids all three of the issues with Nahkola's analysis:

- Geminates are created by the application of the length already attributed by SML rather than by any sort of compensatory lengthening
- The order in which the half-long vowel and elision appeared does not matter
- My model follows a more generative approach, and thus accounts for all words in which the correct context is found


### 3.2.1 Representations of syllable structure

For the representation of syllabic structure and quantity, I assume the model of moraic theory à la Hayes (1989).

Only rhyme constituents are tied to moras $(\mu)$, while onsets are tied directly to the syllable node $(\sigma)$. Short vowels bear one mora and long vowels bear two moras. Codas also bear moras.

An intervocalic geminate serves as both the coda to one syllable and the onset of the following syllable. (10b) shows this, with the so-called "flopped" representation, wherein the consonant is linked to both a mora in the first syllable and directly to the syllable node of the second syllable. The CVCCV word here is Sanna (a proper name). Compare with (10a), showing a representation for the CVCV sana 'word (nominative case)':


The CVCVV word sanaa 'word (partitive case)', then has the representation in (11), where the long vowel bears two moras:


### 3.2.2 Applying SML

Given these representations, we can notate the application of second mora lengthening with a superscript plus sign $(+)$ on the word's second mora. Thus, a CVCV word such as sana in a dialect with SML would be represented as in (12) below. This is implemented phonetically as lengthening on the second short vowel, resulting in the half-long vowel:


### 3.2.3 Creating geminates

I propose a model of primary gemination which uses SML to account for both the diachronic rise and synchronic process of gemination.

Paunonen (1973) has stated that primary gemination originates as phonetic (non-categorical) lengthening on the consonant. Within a few generations of this phonetic "gemination" first appearing, it is reanalyzed as full phonological gemination.

The problem here is again, why does this lengthening originate? What exactly does a long vowel need to be compensated for? The answer is the lengthening of SML itself.

When SML applies to a CVCVV word, what we get is indeed a phonetically overlong vowel in the second syllable, as seen in (13). This is because we have the length attributed by a full two moras, plus the additional length due to SML.


Such an overlong vowel is is not something which needs to be "compensated" for. Instead, the additional length from SML "spills over" onto the onset of the syllable, where it applies the phonetic lengthening to the consonant instead of the vowel which it originally targeted: ${ }^{4}$

[^2]

However, because consonants, being lower in sonority than vowels, are not as stable under gradient lengthening, and so the phonetic length is reanalyzed as a full mora, creating a categorical phonological geminate as in (15). This reanalysis matches up exactly with Paunonen's account.


What motivates the length attributed by SML to "spill over" onto the syllable's onset? When SML applies to a CVCVV word and a an overlong vowel is created, the grammar is faced with a choice: It can either retain the overlong vowel, as though it were going to create a third category of phonological length, or it can do something else with the extra length.

Rather than deleting this extra length after its insertion, the language can instead utilize the singleton consonant in the syllable's onset. This consonant has the phonetic space to host this length, and since it furthermore has phonological space - the language already has a two-way quality distinction for intervocalic consonants - reanalysis to a phonological geminate is to be expected in these circumstances.

### 3.2.4 Implications

Several things should be kept in mind. First of all, there is some optionality as to whether primary gemination applies. Ylitalo (2004:419) notes that primary gemination does not necessarily have to apply: when a native speaker of a dialect with primary gemination is speaking standard Finnish, as they were in her phonetic study, they are fully capable of pronouncing CVCVV words without gemination.

This is because primary gemination creates phonologically categorical geminates, and so the speaker is aware of the process. On the other hand, the half-long vowel caused by SML is a non-categorical process, and so it is present and measurable even in labratory conditions, with careful pronunciation. In other words, a speaker can "switch off" primary gemination, but they cannot "switch off" SML.

The analysis I have presented predicts this, in that it uses a phonetic process which always applies (SML) to create geminates only when they occur in the right context. However, since
the creation of these geminates requires a phonologically categorical shift, it is not something that necessarily must apply. This is reinforced by the fact that the language already has a two-way length contrast, but does not have any contrasts signalled directly by SML.

Nonetheless, we have evidence that a phonetics-phonology interface process (SML), which must reference phonological (moraic) structure, can actually be a motivation to alter that structure categorically when it applies in the right context.

This might be a place to look in future phonological research as a means by which phonetic features manage to "sneak" their way into the phonology of a language. In other words, prosodic structure can be a pathway to phonologization.

## 4 Conclusions

- "Primary gemination" in Finnish dialects represents a truly "phonological" process, in that it occurs relative to categorical prosodic structure (i.e. it does not reference morphology or specific segments)
- It is fed by the application of a phonetic process (SML) within a certain context
- Phonetic processes interacting with prosodic structure can be a pathway to phonologization, particulary in languages like Finnish, where the prosodic system is very structured durationally


## References

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. Linguistic Inquiry 20: 253-306

Kettunen, Lauri. 1940. Suomen Murteet III A. Murrekartasto. Helsinki.
Lehtonen, Jaakko. 1970. Aspects of quantity in Standard Finnish. Jyväskylä: Jyväskylä University.
Nahkola, Kari. 1987. Yleisgeminaatio: Äänteenmuutoksen synty ja vaihteet kielisysteemissä erityisesti Tampereen seudun hämäläismurteiden kannalta. Helsinki: Suomalaisen Kirjallisuuden Seura.
Suomi, Kari, Juhani Toivonen and Riikka Ylitalo. 2008. Finnish Sound Structure: Phonetics, Phonology, Phonotactics, and Prosody. Studia Humoniora Ouluensia 9. Oulu: University of Oulu.

Suomi, Kari and Riikka Ylitalo. 2004. Durational and Tonal Correlates of Accent in Finnish. Journal of Phonetics 32: 35-63.
Paunonen, Heikki. 1973. On the Primary Gemination of Finnish Dialects. Finnisch-Ugrische Forschungen 40: 146-164.

Topintzi, Nina. 2010. Onsets: Suprasegmental and Prosodic Behavior. Cambridge: Cambridge University Press.
Ylitalo, Riikka. 2004. Toisen tavun vokaalin puolipidennyksestä oulunseutulaisten puheessa. Virittäjä 108: 414-422.
Ylitalo, Riikka. 2009. The Realization of Prominence in Three Varieties of Standard Spoken Finnish. Doctoral Thesis. University of Oulu.

Wiik, Kalevi and Ilse Lehiste. 1968. Vowel Quantity in Finnish Disyllabic Words. Congressus Secondus Internationalis Fenno-Ugristarum. Helsinki.


[^0]:    ${ }^{1}$ I owe thanks to Elan Dresher, Keren Rice, Yoonjung Kang, and the Phonetics/Phonology research group at the University of Toronto for numerous comments on this work, as well as to two anonymous abstract reviewers for some additional comments. Any and all mistakes, misgivings, and misspellings are, of course, entirely my own.
    ${ }^{2}$ As in Finnish orthography, long vowels and long consonants are represented with a doubled letter. Non-categorical (phonetic) lengthening, where appropriate, is represented with a length sign (:).

[^1]:    ${ }^{3}$ Judgments were confirmed with a native of Oulu, Finland, the dialect of which shows both primary gemination and the half-long vowel (due to SML).

[^2]:    ${ }^{4}$ One question here might be whether the state of affairs in (14) is one of ambisyllabicity. In terms of the actual phonetics, it is difficult to say how syllabification can be qualified, while in terms of phonological evidence, they do not seem to be truly ambisyllabic, since they are non-phonological long consonants in a language with a true consonantal quantity opposition.

