INTERACTION OF QUANTITY, FOOT STRUCTURE, AND STRESS IN THE 2\textsuperscript{nd}/3\textsuperscript{rd} SYLLABLE SONORANTS OF SOIKKOLA INGRIAN TRISYLLABLES

Natalia Kuznetsova\textsuperscript{1,2}, Irina Brodskaya\textsuperscript{3}, Elena Markus\textsuperscript{4,5}

\textsuperscript{1}Università Cattolica del Sacro Cuore; \textsuperscript{2}Institute for Linguistic Studies, Russian Academy of Sciences; \textsuperscript{3}Queen Margaret University; \textsuperscript{4}University of Tartu; \textsuperscript{5}Institute of Linguistics, Russian Academy of Sciences

natalia.kuznetsova@unicatt.it, i.brodskaya@gmail.com, elmarkus@ut.ee

ABSTRACT

The talk describes interaction between phonological quantity (originally, binary) and the compensatory effects of foot structure in sonorants at the 2\textsuperscript{nd} and the 3\textsuperscript{rd} syllable boundary (C\textsubscript{3}) in the trisyllabic foot of vanishing Soikkola Ingrian (Finnic). Such interaction had been previously observed as a phonetic effect in the 1\textsuperscript{st} syllable segments and as a phonological impact on the 2\textsuperscript{nd} syllable vowels [1]. C\textsubscript{3} sonorants are also affected, both phonologically and phonetically, and their binary length contrast is almost lost. Depending on the foot structure, original C\textsubscript{3} sonorant geminates are either retained, or nearly or completely shortened to singletons. This result is placed within other results on the trisyllabic foot, which appears to be a coherent metrical unit different from both the disyllabic foot and the bifoot trisyllabic word in Soikkola Ingrian.

Keywords: quantity contrast loss, trisyllabic foot, sonorants, temporal compensation, Soikkola Ingrian.

1. INTRODUCTION\textsuperscript{6}

The Soikkola dialect of the Ingrian language (Finnic), spoken at the Russian-Estonian border, manifests the trisyllabic metrical foot, e.g. in markkina ['mark'ina] ‘breakfast’. The trisyllabic foot shows phonetic and phonological differences from both the disyllabic foot, e.g. kurki ['kurgii] ‘crane’, and the combination of a disyllabic and a monosyllabic foot, e.g. markinä ['mur'ni:] ‘breakfast:PRT’. In particular, in a major phonetic study on 22 Soikkola trisyllables [1], the trisyllabic foot showed stronger compensatory effects than the disyllabic foot. Compensatory effects imply a shortening of segments as a function of an increase in the quantity and number of adjacent segments.

In the first syllable segments (V\textsubscript{1} — the 1\textsuperscript{st} syllable vowel, C\textsubscript{2} — a consonant at the 1\textsuperscript{st} and the 2\textsuperscript{nd} syllable boundary, R — an optional sonorant between them, as in markkina), all compensatory effects were purely phonetic. However, in the second syllable unstressed vowel (V\textsubscript{2}), they led to a merger of the binary length contrast in trisyllables, unlike in the disyllabic foot, which still retains the contrast of short and long V\textsubscript{2}.

The present study explores the influence of foot structure on the quantity contrast in C\textsubscript{3} consonants, which follow V\textsubscript{3}. In C\textsubscript{3}, the quantity contrast of consonants in Soikkola Ingrian is ternary, which is an extreme phonological rarity [2–4]: kana ['kana:] ‘hen’ (singleton) — koainä ['kan a:] ‘hen:PRT’ (short geminate) — linnä ['lin:a:] ‘city:PRT’ (long geminate). In C\textsubscript{3}, the contrast is only binary (singleton vs. geminate). C\textsubscript{3} geminates historically correspond to C\textsubscript{2} long geminates. Hence the typical transcriptions of the C\textsubscript{3} length contrast in older Soikkola sources [5] as sūrimä ‘grain’ vs. sūvimma ‘judge:PRT’ (sonorants), hōvata ‘relax:INF’ vs. sivotta ‘cattle’ (stops).

However, geminates in Finnic languages, including Ingrian, are shorter after an unstressed syllable (i.e. as C\textsubscript{3}) than after a stressed syllable (i.e. as C\textsubscript{2}) [6, p. 274, 7]. Moreover, sonorant geminates differ from stops in that they tend to lose gemination and become singletons in the C\textsuperscript{3} position. The shortening of C\textsubscript{3} sonorant geminates has been observed as complete in Estonian and as ongoing in Finnish dialects [6, pp. 273–277] and in Ingrian dialects, including Soikkola Ingrian [8–10], but never experimentally studied.

In our dataset from [1]\textsuperscript{4}, the shortening of C\textsubscript{3} geminate sonorants in Soikkola Ingrian can be studied on the first plural verbal forms with the suffix -mma/-mmä and the nominal forms in allative with the suffix -lle (or -lла/-лла), e.g. sāta-mma ‘get:PRT’, palka-lle ‘salary-ALL’. Original geminate *lː (160 tokens) can be also directly compared to singleton l in forms like kattila ‘cauldron’ (828 tokens). In turn, the forms with original C\textsubscript{3} geminate *mː (988 tokens) do not have any counterparts with singleton m in these data. Datasets on *lː-l and *mː are imbalanced also in other respects (see 3.1). However, the overall variety of the structural types represented in these data is still sufficient for making certain conclusions about the process of shortening of the C\textsubscript{3} sonorant geminates.

We find that C\textsubscript{3} long sonorants are durationally at most short geminates. We confirm an ongoing merger of C\textsubscript{3} singleton and geminate sonorants. This process has a compensatory nature. It is largely regulated by the 1\textsuperscript{st} syllable structure: the more complex the latter, the shorter any C\textsubscript{3}. The C\textsubscript{3} geminates, like the V\textsubscript{2} short vowels, are the longest in the “light” foot with the 1\textsuperscript{st} stressed light syllable (C)V: otanime ['oda:mmä]. This is an effect of stress-induced lengthening, unusually located in the 2\textsuperscript{nd} unstressed syllable of the light foot in Finnic languages and affecting also C\textsubscript{3} duration.
2. DATA AND METHODS

Phrase-final tokens from [1], used for this study, were collected in the Kingisepp district (Russia) in 2014-2016 from five female speakers of north Soikkola Ingrian born in 1929-1936. Speakers were presented with a phrase spoken in Russian and asked to translate it into Ingrian. Mixed effects linear regression models are fitted in R with lme4 on raw durations to study the relations between the structural parameters in the first two syllables of the foot and the duration of C₃ sonorants. The quantity contrast in V₂ has been lost and is excluded from all models. The fixed effects are V₁ length (V vs. V: — short vowel vs. long vowel), C₂ length (C vs. C: vs. C: — singleton vs. short geminate vs. long geminate), and the presence of a sonorant R (yes vs. no) in various interactions. Models also include random by-speaker and by-word intercepts.

3. RESULTS

3.1. Simplification of the datasets on C₃ sonorants

![Figure 1](image1.png)

Figure 1: Mean durations in ms of C₃ (a) *m*: and (b) *l*: grouped by the combinations of V₁ length, C₂ length, and the presence of a sonorant R represented in our data.

<table>
<thead>
<tr>
<th>V₁ R C₂</th>
<th>C₃</th>
<th>C₃</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>m</em>:</td>
<td><em>l</em>:</td>
</tr>
<tr>
<td>short no</td>
<td>105</td>
<td>43</td>
</tr>
<tr>
<td>sing</td>
<td>190</td>
<td>0</td>
</tr>
<tr>
<td>gem</td>
<td>0</td>
<td>137</td>
</tr>
<tr>
<td>yes</td>
<td>0</td>
<td>65</td>
</tr>
<tr>
<td>sing</td>
<td>358</td>
<td>49</td>
</tr>
<tr>
<td>gem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>long no</td>
<td>155</td>
<td>52</td>
</tr>
<tr>
<td>sing</td>
<td>730</td>
<td>0</td>
</tr>
<tr>
<td>gem</td>
<td>0</td>
<td>373</td>
</tr>
<tr>
<td>yes</td>
<td>0</td>
<td>126</td>
</tr>
<tr>
<td>sing</td>
<td>220</td>
<td>0</td>
</tr>
<tr>
<td>gem</td>
<td></td>
<td>143</td>
</tr>
</tbody>
</table>

Table 1: Numbers of tokens per combinations of V₁ length, C₂ length (singleton vs. any geminate), and the presence of R in our data on (a) *m*: and (b) *l:*.  

Figures 1a-b represent the duration of original C₃ geminates *m*: and *l*: and C₃ singletons /l/ as a function of combinations represented in the dataset from [1] of the levels of the three fixed effects in the 1st syllable. Raw means are reported within boxes; the number of tokens is given below each box at the bottom.

These data contain structural gaps, because the data from [1] were not intended for studying C₃. However, this can be partially remediated by conflations along certain parameters. Preliminary statistical exploration of data showed that short geminate and long geminate C₂ did not significantly differ in their impact on C₃ duration in any fitted model. In Tables 1a-b, they are, therefore, conflated into one category of geminates (“gem”), as opposed to C₂ singletons (“sing”).

In the model on *m*: (3.2), also the factor of R is excluded, as it was not significant for those pairs of structures for which it could be studied: V:C–V:RC:, VC:–VRC:. cf. Fig. 1a. This makes the V₁ length * C₂ length matrix of the m-dataset full, cf. Table 1a.

In the l-dataset, C₁ length (singleton l vs. geminate *l:*) is collinear to the unique combinations of the levels of the three structural factors in the 1st syllable. Therefore, the factor of C₁ length itself cannot be included in a model. However, C₁ length can still be explored in the post-hoc pairwise comparisons of means obtained with a model containing the ternary interaction of the factors in the 1st syllable (see 3.3).

3.2. Results of modelling on *m*: (m-dataset)

Fig. 2 shows the results of C₁ *m*: duration modelling according to the following formula: C₁ duration ~ V₁ length * C₂ length + (1 | speaker) + (1 | word), data = 1758 tokens. The interaction is insignificant but is kept in the model, because this helps to study the relationships between the light foot (in bold in Fig. 2) and the heavy foot (any other type of foot).

A post-hoc pairwise comparison of the modelled means in emmeans shows no significant differences in C₃ between the structures with a short vs. a long V₁ when C₂ is a singleton: [ˈoda.mːa] — [ˈsädamːa] (9 ms, SE = 4.6, df = 50, t = 2, p = 0.052; a black line). In both structures, C₃ *m*: still retain gemination.

When C₂ is a geminate, there is also no difference in C₃ between the structures with a short vs. a long V₁ (4 ms, SE = 1.9, df = 53, t = 2, p = 0.06; a red line), but both C₃ *m*: are almost shortened. Their durations of 101 and 106 ms are similar to the duration of word-initial m in our dataset (106 ms, 338 tokens). All other differences in Fig. 2 are significant at p < 0.0001.

The results on *m*: show that C₃ gemination is kept when C₂ is a singleton, but lost when C₂ is a geminate: otaːmna’ [ˈoda.mːa] ‘take:1PL’, sätːaːmna’ [ˈsädamːa] ‘get:1PL’, but [ˈpalkːama] ‘hire:1PL’ (<*palkkämmä), [ˈlakːama] ‘sweep:1PL’ (<*lakkämmä), [ˈkːntːima] ‘turn:1PL’ (<*kännttimmä), [ˈsːtːima] ‘judge:1PL’ (<*sauttimmä), [ˈleːkːama] ‘cut:1PL’ (<*leikkämmä), [ˈhːlikːaːma] ‘abandon:1PL’ (<*hülkkämmä).
3.3. Results of modelling on *l:-l (l-dataset)

Fig. 3 shows the results of C₃ *l:-l duration modelling according to the following formula: C₃ duration ~ V₁ length * C₂ length * Presence of R + (1 | speaker) + (1 | word), data = 988 tokens. The structures which contain original C₃ *l: are shaded dark grey, while the structures with original short l as C₃ are in light grey.

C₃ *l:, similarly to *m:, has the longest duration of 131 ms (marked in bold in Fig. 3; akalle [’aɡa:le] ‘old woman;ALL’). It is significantly different at p < 0.001 from any other type in Fig. 3.

The remaining two types of *l: in Fig. 3 (’palkalle ‘salary:ALL’ and *koikalle ‘bed:ALL’) are shorter: 106 ms and 102 ms. Statistically they do not differ from each other, but differ from all types of originally short *l as C₃, including durationally the nearest types of *l in kattila [’kat:iла] ‘cauldron’ (91 ms) and kāntelō [’kä:nde lā:] (90 ms), although only at p < 0.05 or at p < 0.01 (cf. also with modelled 89 ms of word-initial l; 436 tokens). The difference is anyway too small to be perceived by humans [11], [12].

Gemination of *l: in these two structures can be, therefore, considered nearly or entirely lost: [’palkalle] and [’koikalle].

In the duration of short *l, we observe the same kind of compensatory trend which distinguished the light foot from the heavy foot in case of *m:. The duration of 91 ms of l in kattila with a short V₁ and no sonorant significantly differs from that of l in structures with a long V₁ (78 ms; rūtele [riː.tel’e] ‘argue:IMP’, Mikkula [’mi:kula] ‘St. Nicholas’), with a sonorant (78 ms; kanitele [’kantel’e] ‘kantele’, kūntelē [’kuntel:i] ‘candle’), or both (80 ms; vāntelē [’vāntel:i] ‘turn:3SG’), but only at p<0.05 or at p<0.01. C₃ duration before a long vowel in a bifoot structure kāntelō ‘turn:3SG’ (90 ms; cf. Ftn. ii) is left aside here.

This is clearly just a phonetic shortening of l — a manifestation of how durations in the foot shorten in an inverse relation to the growing number or quantity of segments within the foot, further discussed below.

4. DISCUSSION

4.1. Similarities and differences between *m: and *l:-l

Our data confirm the ongoing shortening of *m: and *l: as C₃, i.e. after an unstressed syllable, in the trisyllabic foot. The earliest data from 1884 [8] did not yet report this process. Data from the 1930s [13] did not attest it for *m: but reported the shortening of *l: in the allative suffix, although only as C₂ in some personal pronouns (e.g. miu-le ~ miu-le ‘me:ALL’, p. 78). However, the data on the allative suffix and the verbal 1º plural from the 1960s [9] and from the 2000s [10] already reflect a wide variation in *m: and *l: length as C₃ in the trisyllabic foot (and as C₂ after a long vowel, like in the personal pronouns above). The present data, although limited, allow us to clarify some mechanisms of the shortening of C₃ sonorants.

Gemination sonorants *m: and *l: undoubtedly retain gemination in the light foot. However, their duration as C₃ (~130 ms) does not reach the duration even of short geminate stops in C₂ (modelled 177 ms in [1]; cf. with 234 ms for long geminate C₂ stops). Shorter duration of C₃ geminates than of C₂ geminates was expected on the basis of earlier Ingrian and Finnish data. Synchronous C₃ sonorant geminates can be phonologically considered as short geminates /C:/.

In this, they might potentially differ from C₂ stops, which impressionistically do not show any signs of phonological shortening in Soikkola Ingrian (in words like sivotta ‘cattle’), but this is yet to be verified on the newly collected data.

In all studied types of heavy foot, *m: and *l: have almost or entirely shortened, apart from the type with *m: and a long V₁ plus singleton C₂ (i.e. “V:C”, sātānma ‘get:1PL’; ~120 ms), where the duration of C₃ *m: does not significantly differ from that of C₃ *m: in the light foot. Still, in the same type of foot, *l: has shortened to ~100 ms (*koikalle > koikale).
To investigate whether the difference in the “V:C” foot between *mː and *lː is systematic or due to some possible gaps in the data, we considered inter-speaker differences in the structures with a singleton C₃ for *mː and *lː. Broadened inter-speaker variability between competing phonological variants is typical of an unfinished sound change [14]. Fig. 4 shows the results of modelling on m- and l- datasets by a formula: C₁ duration ~ First syllable structure * Speaker + (1 | word). In Fig. 4a, all the types from Fig. 1a where *mː: occurs after any type of a geminate are conflated into “*mː_after_gem” in “short_l”. In Fig. 4b, all types from Fig. 1b with a short l are conflated into “short_l” in the light foot and fully short consonants in the conlated types. There is, however, a large inter-speaker variability of *mː and *lː in heavy feet with a singleton C₂ (highlighted by circles in Fig. 4). In particular, in the “V:C” foot, C₃ *mː has actually shortened in 2 out of the 5 speakers.

It appears that C₃ *mː and *lː retain gemination in the light foot, are in a process of shortening in heavy feet with singleton C₂, and are (almost) shortened in other heavy feet. C₃ *mː is a bit more conservative than C₃ *lː. In both cases, we anyway see the effects of the general compensatory trend attested also for other foot segments in [1] (see Introduction). In 4.2, we discuss the relevance of our findings on trisyllabic temporal compensation for stress and foot typology.

Our present data show that this lengthening in the 2nd syllable of the light foot spreads not only to the syllable nucleus (i.e. V₂) but also to the coda. In the structures considered, the coda consists of the first part of C₃ geminates, e.g. o-tam-ma ['oːtamːa]. Stress-related lengthening in the 2nd syllable likely prevents C₃ geminates from phonological shortening.

In the heavy foot, which is in turn characterised by the metrical weakening of the 2nd unstressed syllable, both long V₂ [18], [1] and long C₃ shorten, e.g. *lak-käm-ma > ['lak ama] ‘sweep:1PL’. Both processes are observed as still ongoing in Soikkola Ingrian and complete in more prosodically innovative Estonian.

Table 2 draws a comprehensive picture of the compensatory processes affecting segmental duration and quantity within the trisyllabic foot in Soikkola Ingrian on the basis of this study and findings in [1]. Compensatory processes represent an interaction of the foot structure (first of all, light vs. heavy foot, with finer gradations within heavy feet) and segmental quantity. Word-initial consonant C₁ is not affected, neither the 3rd syllable vowel V₃ is. The latter is always short in the trisyllabic foot (trisyllables with long V₃ are bifoot words). Quantity contrasts in the 1st syllable are influenced by compensatory shortening only phonetically. The duration and length of V₂ and C₃ are affected both phonetically and phonologically.

The concept of ternary rhythm is contradictory in modern formal phonology. Some deny its relevance altogether and re-analyse known cases of trisyllabic feet through regular di- and monosyllabic foot types [19]. Others propose a recursive (two-layered) foot, where a “minimal” disyllabic foot is embedded into a “maximal” disyllabic foot [20], [21].

Our phonological and phonetic data on Soikkola Ingrian rather suggest that the trisyllabic foot is a coherent metrical whole in this variety. It is a type of a rhythmic group on its own, different from both the disyllabic foot and the bifoot structure. Especially our present data on C₃ geminate sonorants, which are at the boundary of the 2nd and the 3rd syllable but are still compensatorily affected by the structure of preceding syllables, might pose challenges for the above-mentioned formal accounts on ternary rhythm.

---

**Figure 4:** Modelled durations of (a) *mː: (left) and (b) *lː-l (right) as C₃ with a focus on structures with singleton C₂

**4.2. Significance of results for stress and foot typology**

Our most robust finding as regards the C₃ geminate sonorants is their clear preservation in the light foot, as opposed to their shortening in the heavy foot. This fact can be also related to a phonetic lengthening of the preceding short V₂ in the light foot (modelled 127 ms — durationally the longest V₂ type in [1]).

This cross-linguistically untypical phenomenon of lengthening of unstressed V₂ rather than of stressed V₁ in the light foot in Finnic languages has been long known as a “half-long vowel” [15], [16]. For Standard Finnish, the “half-long vowel” has been explicitly considered as a manifestation of stress [17].

**Table 1:** Prosodic structure of the Soikkola Ingrian trisyllabic foot, as observed in [1] and in this study.
7. REFERENCES


Author contributions: The design and the background of the study were developed by NK; the data were collected by NK, IB, and EM, segmented by IB with a help of EM, extracted and processed by IB. NK is responsible for the statistical analysis and visualisation of the data presented in this paper. The draft text was written by NK, revised and edited by NK with a help of EM. NK is the corresponding author and responsible for the definitive version of the text.

The full dataset, including the data used in this paper, will appear at https://osf.io/rwb3m/.

This type with a final Vː was the only bifoot structure among our original 22 trisyllabic types, to cover the structure “VːRC” absent from the trisyllabic feet with C2 stops; cf. Footnote (c) to Appendix 1 in [1].