

<https://helda.helsinki.fi>

---

## Metric Variation in the Finnic Runosong Tradition : A Rough Computational Analysis of the Multilingual Corpus

Sarv, Mari

ICL CAS  
2021

---

Sarv, M., Kallio, K., Janicki, M. M. & Mäkelä, E. 2021, Metric Variation in the Finnic Runosong Tradition : A Rough Computational Analysis of the Multilingual Corpus . in P. Pöyry, P. Plecháč, R. Kolár, A.-S. Bories & J. Xíha (eds), Tackling the Toolkit : Proceedings of the 11th International Conference on Computational Literary Studies . ICL CAS, Prague, pp. 131-150 . <http://hdl.handle.net/10138/339829>

---

<http://hdl.handle.net/10138/339829>

<https://doi.org/10.51305/ICL.CZ.9788076580336.09>

---

cc\_by  
publishedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

# Metric Variation in the Finnic Runosong Tradition: A Rough Computational Analysis of the Multilingual Corpus

Mari Sarv

Estonian Literary Museum, Estonia

[mari@folklore.ee](mailto:mari@folklore.ee)

 0000-0001-5309-2357

Kati Kallio

Finnish Literature Society, Finland

[kati.kallio@finlit.fi](mailto:kati.kallio@finlit.fi)

 0000-0002-3673-1409

Maciej Janicki

HELDIG, University of Helsinki, Finland

[maciej.janicki@helsinki.fi](mailto:maciej.janicki@helsinki.fi)

 0000-0003-3981-8021

Eetu Mäkelä

HELDIG, University of Helsinki, Finland

[eetu.makela@helsinki.fi](mailto:eetu.makela@helsinki.fi)

 0000-0002-8366-8414

## Abstract

This article represents a first step in the corpus-based study of metric variation in Finnic runosong, a poetic tradition shared by several Finnic peoples and documented extensively in the 19th and 20th centuries. Runosong metre has generally been assumed to be a syllabic tetrametric trochee with specific rules about the placement of stressed syllables according to their quantity: long stressed syllables occupy the strong positions in the trochaic schema while short stressed syllables appear in the weak positions. Recent studies by Mari Sarv (2008, 2015, 2019) of Estonian runosong metre have shown, however, that due to linguistic changes, it has gradually lost its quantitative properties and acquired the features of accentual metre.

Using computational methods, this study aims to give a preliminary overview of the extent of metric variation on the quantitative-accentual scale across the entire Finnic runosong area. After an approximate syllabification, we apply two separate indirect methods for estimating variation. These appear to generate coherent results: quantitative runosong metre dominates in the north-east and has gradually been replaced by accentual runosong metre towards the south-west. Subsequent studies should verify these results through more precise and detailed investigations.

## 1 Introduction

The present paper forms the first part of a larger study which aims to describe the metric features and variations in the corpus of a poetic tradition that is documented across almost the whole Finnic area. The entire corpus comprises

over 240,000 runosong texts in Estonian, Finnish, Karelian, Ingrian (Izhorian) and Votic that were mainly recorded during the 19th and 20th centuries. We seek to answer such questions as, What are the characteristics of runosong metre? What features appear across the corpus? And how do they relate to poetic genres and regional singing cultures? Statistical analysis of the corpus may allow us to uncover trends and connections not detected in previous research, which was limited to selected regions, sub-genres and singers. The current article offers a preliminary estimation of the extent of metric variation within the corpus based on an approximate syllabification.

Oral poetry only exists in fleeting performances and memory structures. The texts that we have are more or less accurate transcriptions of individual performances. Neither the poems nor any individual verses have a single original or authoritative form; rather they developed from performance to performance and from singer to singer. Typically singers of these works have a number of options for varying the performance of their poems and lyrics. Albert Lord (1960) notes that in many oral traditions, singers tend not to use poems learned verbatim by heart; instead they create a poem during its performance based on their knowledge of traditional storylines, formulae (metrically and poetically motivated collocations), poetic language and the needs of each situation.

The creation and transmission process peculiar to folklore inevitably produces variation across different aspects of folkloric expression. This also applies to the corpus under investigation and needs to be kept in mind with respect to this research. The properties of the texts vary based on the regional peculiarities of their content, performance, usage, poetics and language. Their metre also varies both as a standalone poetic feature and as a result of other layers of variation. Regional and genre-based variations in rhythmic melodic structure, for example, place different restrictions on metric structure. Given all of its linguistic and poetic variation, the current corpus, thus, poses serious challenges for a large-scale computational analysis. In addition, there are biases in the geographical and typological distribution of this material that stem from regional traditions and the collection history. As such, any quantitative results require careful interpretation and contextualisation.

## 2 The Finnic Runosong Tradition and Its Metric Properties

Most Finnic languages (see [Figure 1](#)) share a specific poetic-musical tradition that is characterised by stichic diction (no stanzas or rhymes), interdependent use of alliteration and parallelism and a syllabic metre with a trochaic core.<sup>1</sup> In the English-language scholarship, this tradition has various names of which the most common are Finnic oral poetry, Kalevalaic poetry and runosong although researchers have also noted the shortcomings of these various terms ([Kallio et al. 2017](#)). What is distinctive about this poetic form is its use across a wide variety of sung genres (e.g., epic, lyric and ritual songs, occasional songs, charms

<sup>1</sup> This tradition is, however, unknown among Livonians and Vepsians. Audio recordings of traditional runosong together with original texts and English translations can be accessed online; see, for example, the samples in [Tampere et al. \(2016\)](#) and [Oras \(2015\)](#).



Figure 1: The main dialects of the Finnic languages at the beginning of the 20th century (Grünthal forthcoming)

and improvisations) and also to some degree in colloquial speech and spoken word genres such as proverbs, sayings and riddles (Krikmann 1997; Kuusi 1994; Tampere et al. 2017). Alongside this prevalent poetic system, other less common poetic structures in various genres (e.g., laments and children's songs) also evolved in this region. Sometimes, as in the case of bagpipe songs, they were influenced by other cultures. Eventually, rhymed stanzaic songs with Russian, German and Swedish influences prevailed in the popular poetic expression of the second half of 19th century. Today these traditions continue in various forms.

The earliest preserved examples of runosongs were transcribed in the 16th and 17th centuries, and the majority of related texts were recorded during the 19th and early 20th centuries. The beginning of this folklore collection was tied to the awakening of national consciousness in Finland and Estonia, then both part of the Russian empire – and the need for both these small states to create a national culture and history (see, e.g., Anttonen 2005; Tarkka et al. 2018). The nationalist call for collection generated extremely rich and voluminous

collections. From 1904 onwards, songs were also recorded. The first recordings were made on phonographs but later other audio recording techniques evolved and were used (see, e.g., Oras 2008; Huttu-Hiltunen 2008).

The metre of Finnic runosongs varies by region but is generally thought to derive from a common original form: this is a syllabic metre with a trochaic core that has four stresses (and feet) and follows specific rules about the quantity of stressed syllables (in Finnic languages, the stress falls, as a rule, on the first syllable of a word). This metre is believed to have been established during the Proto-Finnic linguistic phase when the prosodic system of the language proved itself suitable for the evolution of syllabic metre (Korhonen 1994). This was in the first millennium BCE, a period when according to contemporary archaeogenetic research, the Finno-Ugric genomic component arrived in the area around the Baltic Sea (Saag et al. 2019).<sup>2</sup> The poetic system of this tradition is thought to have developed from (1) earlier poetic forms;<sup>3</sup> (2) changes in the prosodic structure of the language; and/or (3) contact between different cultures in this era when Proto-Finnic had not yet divided into many separate languages (Frog 2019; Korhonen 1994; Sarv 2019).

Runosong metre was first described vaguely by 17th-century scholars (Sarajas 1956). In his work *De poesi Fennica*, Henrik Gabriel Porthan, the most notable early researcher of runosong, mentioned the specific role of trisyllabic words in the form's rhythm. Ultimately, however, he echoed his predecessors' view that syllabic quantity did not play a role in runosong metre (Porthan 1983, p. 41). The quantitative basis of this metre was discovered by Arvid Genetz, who in 1881 formulated the "quantity rules" of *Kalevala* metre, relying mainly on Karelian runosongs (Genetz 1884 [1881]).<sup>4</sup> These rules describe the placement of stressed syllables according to their quantity: except in the case of the first two line positions, long stressed syllables occupy the strong positions in the trochaic schema while short stressed syllables fill the weak positions. These rules, thus, describe and explain the very essence of the peculiar rhythmic alterations of runosongs, something that researchers had been pondering for decades.

Below we provide an example. This is a wedding song from northern Estonia in quantitative runosong metre:

Pei-u-ke-ne, <b>poi</b> -si-ke-ne, (44)	Bridegroom, dear boy,
ei <b>mi</b> -na <b>ra</b> -ha-ta <b>lau</b> -la, (1232)	I'm not singing without payment,
kul-la-ta keelt en <b>ku</b> -lu-ta, (3113)	not using my tongue without gold,
keelt en <b>pek</b> -sa <b>pen</b> -nin-gi-ta! (1124)	not beating my tongue without a penny!
Mi-na <b>lõ</b> (p)-pe-tan <b>lu</b> -gu-da, (233)	I'm going to make up my song,
si-na <b>kae</b> -va <b>kar</b> -man-tu-da, (224)	you'd better dig into your pocket,
ra-ha <b>tas</b> -ku-da <b>ta</b> -o-ta! (233)	reach for your wallet!

<sup>2</sup> Other scholars have made similar assumptions about the general period of runosong's genesis based on different information; see, for example, Leino (1986, p. 140), Rüütel (1998) and Frog (2019).

<sup>3</sup> Speculation about the common Uralic or Finno-Ugric poetic and metric forms that might have preceded runosong falls outside the scope of this paper. For more information on this topic, however, see, for example, Korhonen (1994), Helimski (1998) and Frog (2019).

<sup>4</sup> Elias Lönnrot had also used Karelian runosong when composing his epic, *Kalevala*.

These lyrics have been syllabified while we have marked the stressed syllable/s in the polysyllabic words in metrically relevant positions in bold. After each line of verse, its syllabic structure is also given with numbers that indicate the syllabic composition of the line's wording. The line structure 233, for example, consists of a disyllabic and two successive trisyllabic words.

E, StK 8, 86 (2) < Lügänuše – Mihkel Jürna, stud. < auntie of Karja, 85 years old (1921).

Since Genetz's discovery, these quantity rules have been absorbed into the general descriptions of runosong metre in Finnish and Estonian scholarly as well as educational materials. They have also been assumed to apply to runosong more generally without any further verification of this claim. Subsequent scholarly research on this metre has either (1) considered deviations to be inconsistencies peculiar to folk metre (Anderson 1935), (2) been based on parts of the *Kalevala* itself or on songs from the most *Kalevala*-metric areas, genres or singers (see, e.g., Saarinen 2018; Sadeniemi 1951) or (3) discussed the origins and features of *Kalevala* metre as a model without analysing specific texts (see, e.g., Korhonen 1994; Kuusi and Tedre 1979).

At the same time, systematic deviations of folksong metre from the *Kalevala* model have sometimes been noted. Estonian scholars have, thus, sometimes introduced the rules of *Kalevala* metre with the caveat that they are "statistical", i.e. they apply to the majority of verse lines, without further specifying the percentage of lines or texts that this affects (e.g. Viidalepp 1959, p. 121). On the one hand, regional differences have been noted, especially in southern Estonia, while on the other, these deviations have been attributed to foreign influences and the disintegration and extinction of the tradition (Viidalepp 1959, p. 126). The only researcher to question the relevance of *Kalevala* metre to Estonian runosong was the literary scholar Jaak Põldmäe, who raised the issue in his *Estonian Metrics* (1978). His approach, however, proceeded from a tradition of literary analysis and assumed consistent rules throughout a song. As such, it neglected the consequences of the transmission process and the variations characteristic of folklore.

With some exceptions, most Finnish research has concentrated on Karelian poems, i.e. on the most regular and *Kalevala*-metric part of the tradition. Pentti Leino (2002 [1975]) compared the metre of three groups (19th-century folk poets, earlier scholarly poets who used versions of traditional metre and some 20th-century Karelian singers) to *Kalevala* metre, noting interesting variations both in the syntax and metre of these poems. He also observed that in southwest Finland, the quantity rules about short stressed syllables were not as strict (see Laitinen 2006, p. 38; Leino 1994, p. 71). Matti Kuusi (1983, pp. 184–187) analysed the repertoire of one Ingrian Finnish singer, Maria Luukka from western Ingria and pointed out that researchers he needed to standardise some of her verses and that in the case of the shorter verses, it was impossible to know how they were performed (and thus, how singers had traditionally interpreted these verses in relation to the metrical system; see Oras 2010; Kallio 2021). Similarly, Petri Lauerma (2001, 2004) examined the metre and language used by the famous Izhorian singer Larin Paraske from northern Ingria. In this quite exceptional case, recorders had transcribed both the dictated and sung versions

of the same poems and also noted Paraske's explanations. Lauerma concluded that the words in the dictated lines, and especially the line endings, often took shorter forms that came closer to the contemporary spoken language, while the sung performance used more regular lines, i.e. with all the verse positions filled.

The question of “metric dialects” of runosong was first raised by Kuusi (1983, pp. 188–190, 1994, p. 55), who also expressed a wish to carry out further research. As part of this project, he proposed that the Finnic area could be divided metrically into two regions: (1) the *Kalevala*-metric zone, including Karelia, eastern Finland and Ingria and (2) the late-*Kalevala*-metric zone of western Finland, southern Karelia, eastern Ladoga Karelia and Estonia where, according to Kuusi, the tendency to adopt accentual metre had developed in the 16th century (Kuusi and Tedre 1979, p. 70). This suggestion was not elaborated further, however, since verification would have required an immense amount of repetitive and time-consuming analysis in the pre-computer era.

In fact, the thorough study of the variability of runosong metre only became feasible with the aid of computing technology (e.g. Sarv 2008, 2015, 2019). In the case of the Estonian material, such study has demonstrated that (1) there are clear and significant regional differences in Estonian runosong metre and (2) Estonian runosong metre represents a transitional stage between *Kalevala* metre, where the quantity of stressed syllables is a distinctive feature, and an accentual metre that relies on the placement of stressed syllables without any additional quantitative rules. The main difference between the two metrical variants lies in the placement of short stressed syllables: in quantitative runosong metre, these occupy the weak positions in the trochaic schema, while in accentual runosong metre, there is no difference quantitatively, i.e. just like their long counterparts, the short stressed syllables may be placed in strong positions (see Table 1). Throughout Estonia, both variations were used simultaneously although the proportion of lines following each metre varied significantly by region.

We still, however, lack even a general picture of the variability of runosong metre in the other Finnic regions. We have neither the slightest idea of the extent of metric variation, nor any sense of whether or how this might relate to linguistic borders. The current study is the first enquiry in the research project “Formulaic Intertextuality, Thematic Networks and Poetic Variation across Regional Cultures of Finnic Oral Poetry FILTER” (funded by the Academy of Finland), which aims to understand the metric properties of this vast corpus of material. The main goal of this first step is to obtain a rough idea of the extent of geographic variation of runosong metre using the data at our disposal and automatic analysis only. We are, of course, aware that such an approach is not equipped to handle details or special cases. It may, however, offer ideas and hints about how to plan a more substantial enquiry. At the same time, it may reveal how biases in the data are reflected in the automatic analysis.

Trochaic rhythm – intermittent <b>strong (S)</b> and <b>weak (W)</b> verse positions: SW/SW/SW/SW	
QUANTITATIVE RUNOSONG METRE	ACCENTUAL RUNOSONG METRE
<b>Long</b> stressed syllables of polysyllabic words are placed in <b>strong</b> line positions in both metrical variants:	
näin mie / <b>un</b> -ta / <b>moa</b> -tes-/sa-ni (verse structure 1124) “I dreamt a dream while sleeping”	
<b>Short</b> stressed syllables are placed in <b>weak</b> line positions in <b>quantitative</b> runosong metre:	<b>Short</b> stressed syllables are placed in <b>strong</b> line positions in <b>accentual</b> runosong metre:
jo <b>tu</b> -li <b>sy</b> -vä <b>sy</b> -ky-sy (1223) “a deep autumn set in”	<b>o</b> -leks / <b>mi</b> -nu / <b>o</b> -le-mi-ne (224) “if it were up to me”
joo-tis / lau-li-/ku <b>ka</b> -bas-ta (233) “let me drink from a singer’s cup”	Kad-ri / var-vad / <b>va</b> -lu-/ta-vad (224) “Kadri’s toes are aching”

Table 1: Principles for the placement of stressed syllables in two concurrent runosong metres: (1) quantitative runosong metre (*Kalevala* metre) and (2) accentual runosong metre. In the examples, the stressed syllables relevant for detecting the metrical variant (i.e. quantitative or accentual) are highlighted in bold.

### 3 Research Material

The material in the current study consists of runosong texts and metadata from two major runosong databases: (1) the Estonian runosong database (Oras et al. 2003–2021) and (2) the SKVR database of Karelian, Ingrian and Finnish runosongs (SKVR 2021).<sup>5</sup> These databases were developed in parallel and have broadly the same structure in which texts and original metadata are formatted in extended markup language (XML) and are combined with tables about the unified metadata as well as other classification information. The databases have been united in a joint system for research purposes within the FILTER project.

Despite the general coherence of the two databases, there are some slight differences in their composition principles and history that should be mentioned. The Estonian runosong database was formed directly from the archival collections at the Estonian Folklore Archives, which consist mainly of Estonian and Seto material and to some extent also Ingrian and Votic songs. The key features of this database are as follows: (1) it is not complete but contains approximately two thirds of the runosongs in the Estonian Folklore Archives; these are the oldest collections and the database is continuously being supplemented with additional texts from the newer collections (the data for the current study con-

<sup>5</sup> Concerning the background and development of these databases, see Sarv 2020; Harvilahti 2013 and Klemettinen 2006.



sisted of the 100,034 texts that had been added to the database by June 2020); (2) the texts in the database have not been fully classified; they are provided with preliminary classification data from the archival indexes, and the revision of the system is ongoing; (3) although the main body of the database consists of runosongs, it also contains some material from other poetic folklore genres and hybrid poetic forms; in the current stage of classification, these cannot be fully distinguished from runosongs based on the metadata; and (4) the database contains an orthographically unified version of the song texts. The SKVR database, in contrast, arose from an academic edition of runosong texts, *The Ancient Poems of Finnish People (1908–1997)*, which drew on the folklore collections of the Finnish Literary Society. It contains a total of 89,247 texts of Karelian, Finnish, Ingrian, Votic and Ingrian Finnish runosongs. For our analyses, the following attributes of the SKVR database should be kept in mind: (1) it includes all of the poems listed in the above publication but not all runosong texts in the Finnish archives; (2) it contains texts that are fully classified based on years of effort by several researchers; (3) it consists mainly of runosong texts; and (4) it is orthographically varied and, depending on the language or dialect in question, uses special marks to reflect pronunciation (for search purposes, the SKVR database also offers an automatically standardised or simplified version of texts where these special marks are removed and specific letters are replaced; see [skvr.fi/ohje](http://skvr.fi/ohje)).

The corpus is uneven in many ways. First, there are considerable differences in the distribution of subgenres of runosong. In Estonia, for example, traditional wedding ceremonies were occasions accompanied and structured by runosongs, and wedding songs probably form the majority of Estonian runosongs. In contrast, in Finland, wedding songs are almost unknown. In Finland, Karelia and Ingria, long and narrative incantations represent a considerable part of the runosong collection, while in Estonia, incantations in runosong form comprise only a few types and they are considerably shorter in nature. Although collectors recorded all kinds of poems in these archives, they preferred metrically and poetically coherent epic or mythological works. Some kinds of songs tended to be considered unsuitable for recording, particularly if they were more improvisational or covered sexual topics. Regions such as Viena Karelia and Setomaa, where the traditions were more active and versatile during the collection period, were prioritised and surveyed most closely. The extent of the runosong collection, thus, differs substantially across regions.

The runosong texts in the corpus were transcribed from several Finnic languages that are not in themselves homogenous but have dialects and subdialects. As well as exhibiting purely linguistic variation, runosongs often conserved archaic and usually longer word forms. In some regions, runosong language is a specific archaic idiom that is structurally different from the spoken language based on its prosodic and morphological features. The peculiarities of texts in dialects that lack a consistent literary standard have sometimes motivated collectors to invent their own writing systems to record phenomena and sounds not present in literary Finnish or Estonian. Some words in the research corpus may have hundreds of different forms due to dialectal variation, archaic runosong idioms, orthographic inconsistencies and morphologic variants. There are no

dictionaries or automatic tools for lemmatisation or grammatical analysis of this variability; even the simplest syllabification rules may apply differently across spoken languages as well as across runosong forms.

## 4 Methodological Considerations

For the computational analysis of a syllabic metre with subsequent quantity regulation, it would seem natural to use line syllabification as a starting point. This would involve detecting the stress and quantity of the syllables to see how they map onto the octapositional verse lines. However, such an automatic analysis of our corpus is complicated for several reasons:

1. Although runosong metre is syllabic in nature, regionally different replacement rules apply; for example, one position is often filled by two light syllables while in some regions, a long or overlong syllable may be stretched over two line positions.
2. Syllables with long vowels, and especially diphthongs that historically derive from two syllables, behave flexibly: they may either fill one position in the verse schema or be divided between two positions. In addition, there are no strict syllabification rules for this kind of long vowel that could serve as a basis for automatic syllabification.
3. In the case of compounds, the stress on the first syllable of the second component is also relevant to the metre but there is no quick way to divide compounds automatically.
4. Depending on the region, the eighth line position sometimes remains unfilled or the last two line positions may be filled by one long syllable, which in historical linguistic practice would have been two syllables.

Syllabification (and performance) practices vary by genre, region and individual. Singers may have interpreted verse rhythms based on different linguistic layers or made choices about lyrics based on different interpretations of word structures. A combination of these factors complicates the automatic syllabification process. This affects the detection of which syllables are stressed and whether they should be considered short or long as well as the final mapping of the syllables onto positions in the verse schema.

In addition, there are deviations that affect only minor parts of the corpus:

1. Although most works in the runosong tradition used a performance style in which a structural melodic note corresponded to a verse position (generally a syllable), specific melodies and regional performance styles may have required repetition or additional structural syllables. This may, in turn, have affected a song text and occasionally its metre. Performance factors may also have influenced how singers created songs and verse. The performance style could, for instance, affect the number of syllables in words and lines by causing singers to make verses fit into rhythmic song structures in various ways.

2. There are different linguistic and orthographic variants and peculiarities across the corpus that may also influence the syllabification rules in some cases (in addition, some collectors tended to “translate” songs in local dialects into standard Finnish or Estonian).
3. Recorders sometimes used abbreviations and incomplete wording; singers also used different and less metrical word forms when they dictated a song as opposed to when they sang it (see, e.g., [Lauerma 2004](#)); some spoken poetic genres like charms and songs for children tended to take less regular verse forms.

In sum, all of these circumstances pose a substantial obstacle for anyone attempting a quick but sound automatic analysis of the metre throughout the corpus. This led us to look for an indirect approach that might provide a preliminary idea of the extent of metric variation in the area associated with runosong.

In proposing a method to indirectly estimate metrical variation across the corpus, we relied on the results of an analysis of Estonian runosong metre by Mari Sarv (2008) that was based on a sample of around 50,000 verse lines (equal samples of 500 lines from each of the Estonian parishes). This study had revealed that the regional variability of metre clearly overrides any metrical differences among the main song genres (lyric songs, lyroepic songs and ritual songs), variability due to recording notation methods or differences among individual singers. The metre of Estonian runosongs turned out to follow a pattern of clear geographical variation. Second, this study managed to demonstrate that the features of verse metre depended on dialectal prosody ([Sarv 2008, 2019](#)). As many past researchers have noted, the metres widely used in a language need to conform with its prosodic system, especially in the case of indigenous folk metres that develop over extended periods together with the language (see, e.g., [Ross et al. 2001](#), p. 3; [Lotman 1998](#), pp. 1853, 1858–1859; [Jakobson 1979](#), pp. 148–150).

In the case of Estonian runosong, where metrical changes reflected substantial changes in the language, the clearest and most straightforward linguistic correlate with metre was the average length of words in song texts: in parishes where the majority of lines reflected typical runosong quantitative verse, the average word length in syllables tended to be longer. In contrast, in parishes where most lines conformed with typical runosong accentual metre, the average word length was shorter. The correlation between the percentage of typical quantitative lines and word length across 104 parishes was 0.83. The reasons for this correlation can be traced to a set of major historical changes in the Estonian language that systematically decreased the number of syllables in words. To adapt to these changes, the metre took on a new form better suited to the prosodic system of the revised language. These linguistic changes did not take place all at once or across all of Estonia but rather gradually and with different effects on Estonian dialects, and thus, also on runosong metre.

Concerning biases, Sarv’s study showed that on average, more runosongs had been recorded from parishes with a higher percentage of runosong lines typical of quantitative metre (the correlation between the percentage of lines typical of quantitative runosong metre in a parish and the number of songs from

the same parish in the Estonian runosong database was  $r = 0.49$ ). On the one hand, this may point to a connection between metrical changes and the fading of tradition (the active tradition in some parishes kept the metre from changing abruptly). On the other, we may assume that the collection process relied on the material available at a time when the tradition was gradually vanishing: it would have been hard to find singers in places where the tradition was about to disappear and easier to collect songs in those with an active practice. Collectors sometimes also preferred to record songs from regions where songs were longer and followed quantitative metre because these were seen as the characteristics of “true” runosong.

Based on Sarv’s findings, we hypothesised that the average word length in the runosongs recorded in a parish might reveal the extent of runosong metric variation on the quantitative-accentual scale. We did not, however, have any direct means to automatically test this hypothesis and therefore needed to take an alternative approach: since the most common line structures typical of quantitative runosong metre (233, 323, 332) have no regular match in accentual runosong metre, we used their occurrence as a complementary indicator of metric variation.

## 5 Analysis and Results

The two aims of this preliminary research were (1) to calculate the average number of syllables in words in the current corpus by location and region and (2) to determine the percentage of line structures typical of quantitative runosong metre. In order to achieve these goals, we first needed, however, to syllabify the corpus. For the approximate syllabification of the whole corpus, we applied Finnish syllabification rules irrespective of differences in language, dialect or orthography. We were aware that this rough method might result in occasional mis-syllabification, in particular concerning: (1) diphthongs and long vowels and (2) compounds. Treating compounds as one word rather than two or three words systematically increases the average number of syllables and, to some degree, distorts line structure percentages. We assumed, however, that these occasional and systemic differences would affect regional texts fairly evenly and so would not seriously modify general trends.

In order to test our assumptions, we first compared the average word lengths obtained from 1) the rough syllabification of the whole Estonian part of the corpus and 2) the results of Sarv’s 2008 study based on samples of 500 lines per parish. In the latter, the compounds had been divided and the automatic syllabification of diphthongs was re-checked manually. It must be reiterated that the corpus of the current study, which includes far more varied material, was somewhat different from Sarv’s research sample where certain formal criteria served to restrict the material to only typical runosongs.

As expected, the results of this test demonstrated a clear positive correlation ( $r = 0.80$ ) between the average word length in syllables based on the current rough syllabification of the corpus, and the average word length found in the previous study (Figure 2). Closer examination of the divergence of the results

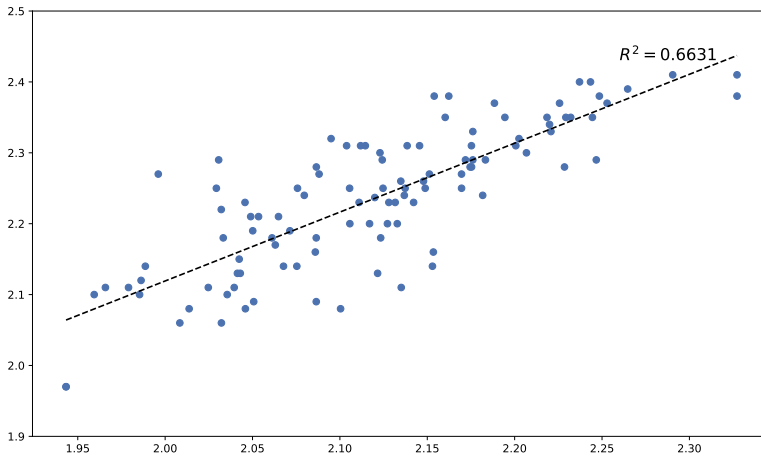


Figure 2: A comparison of the average word length in syllables based on 1) the rough syllabification of the Estonian part of the corpus of the current study (*y*-axis) and 2) Mari Sarv's 2008 study of Estonian runosongs (*x*-axis) with a linear regression line. Each dot represents a parish in Estonia.

from the linear regression line (or the predicted results) revealed that in general the results matched the trendline better in parishes where more material was available; greater differences were recorded in parishes where there was less material (fewer words) in the corpus (Figure 3). There appeared to be no clear regional difference in this distribution. The correlation between the percentage of lines typical of quantitative metre in the previous study and the average length of words in syllables was almost the same for the rough corpus-based syllabification ( $r = 0.84$ ) and the re-checked syllabification from the previous study ( $r = 0.83$ ). These results, thus, suggested that our rough syllabification was a feasible method at least for detecting the average word length in syllables, which could then be used to assess the regional variation of runosong metre on the quantitative-accentual scale.

The maps below present the results of two independent surveys in our study and give a preliminary estimate of the extent and spread of metric variation across the whole runosong-associated region.

The average word length in syllables for the entire corpus is depicted in Figure 4 (per parish) and Figure 5 (per county or dialect area).<sup>6</sup> In Figure 4, the results are shown in greater geographical detail. The less granular view in Figure 5 relies on the regional division in relevant databases. This view counterbalances the extreme results that tend to occur in cases where only a few recordings are available from a parish.

Figures 6 and 7 show the total percentage in the corpus of the most common line structures typical of quantitative runosong metre (233, 323, 332).

<sup>6</sup> In the maps on the following pages, parishes/regions are divided into four classes of approximately equal total area. The colours in these maps represent different intervals for the given variable.

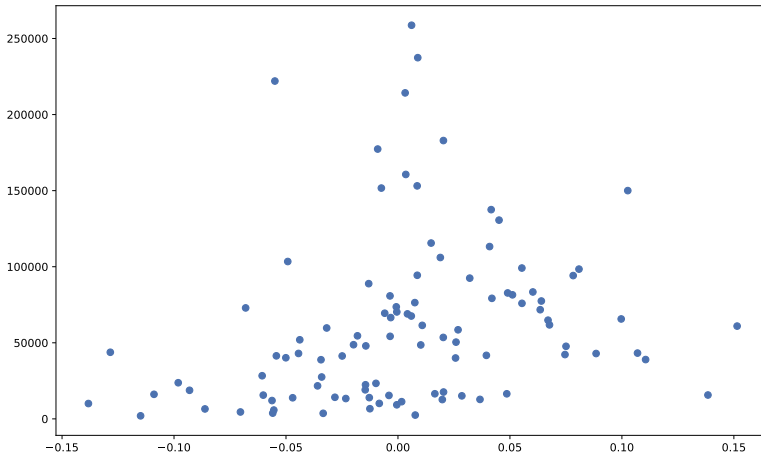


Figure 3: The deviation of the results of our corpus-based syllabification from the predicted average word length in syllables (trendline from Figure 2) is shown on the  $x$ -axis. The number of words in the corpus from the given parish appears on the  $y$ -axis. Each dot represents a parish in Estonia.

The results of the two surveys are generally congruent with one another and reflect a clear geographical logic: the indicators of quantitative runosong metre, i.e. the average word length and percentage of specific line structures, decrease gradually from the north-east (Karelia) to the south-west (Estonia). The somewhat scattered picture in the results at parish level (Maps 4 and 6), especially in Finland, is probably a result of scant collection data, which tends to represent a specific selection rather than the “average runosong” of the parish. As we have seen (Figure 3), where less material is available, the outcome is less balanced, and this leads to more diffuse results.

The correlation between the parish-level results for the two surveys is 0.61. However if we exclude from this calculation the 109 parishes where fewer than 100 verse lines were collected, the correlation rises to 0.76. Across the counties/dialect areas, the correlation between the two surveys is 0.79. The correlation between these two features suggests that longer words are required to form the lines typical of quantitative runosong metre. In this respect, the deviations from the expected values based on a linear regression line can be seen on Figure 8. We see that in Karelia, eastern Finland and Ingria, the percentage of lines typical of quantitative runosong metre is higher than expected based on the average word length in syllables. On the other hand, there are regions where the traditional use of lines typical of quantitative metre has declined despite the length of the words. In these regions, the Finnish population has mixed with either the Swedish or Sámi population and this may have had an impact on the runosong tradition. (Sarv (2011) describes a comparable situation in the Estonian borderlands.) The deviations in either direction from expected values probably relate to the tension between the endurance and disappearance of the runosong tradition in these areas.

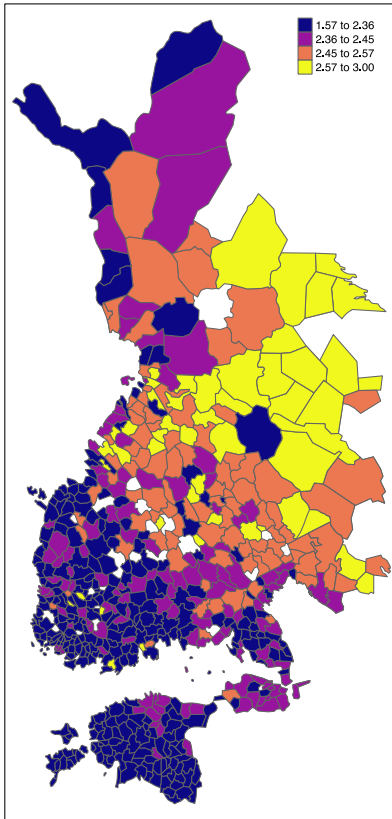


Figure 4: Average word length in syllables in the runosong texts for each parish in the corpus

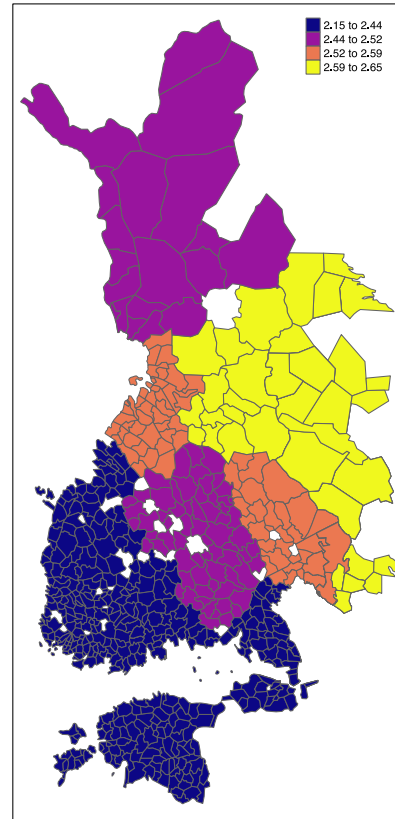


Figure 5: Average word length in syllables in the runosong texts for each region (county or dialect area) in the corpus

Based on the results of this study, we may assume the following:

1. the metre of runosongs is generally more quantitative in Karelia and eastern Finland and less quantitative in Ingria, western Finland and Estonia (this generally conforms with the earlier proposal by Matti Kuusi except that Ingria belongs to the less quantitative region in the current results) and
2. changes in metric features depend on changes in linguistic/dialectal prosody (as is most clearly shown in the shortening of words in our study). Such changes do not necessarily follow language or dialect borders (cf. Sarv 2008, 2019).

The previous Estonian study (Sarv 2008) showed that lines typical of quantitative runosong metre dominated the tradition across half of Estonia, while in the other half, accentual metre prevailed. The more quantitative metre-defined

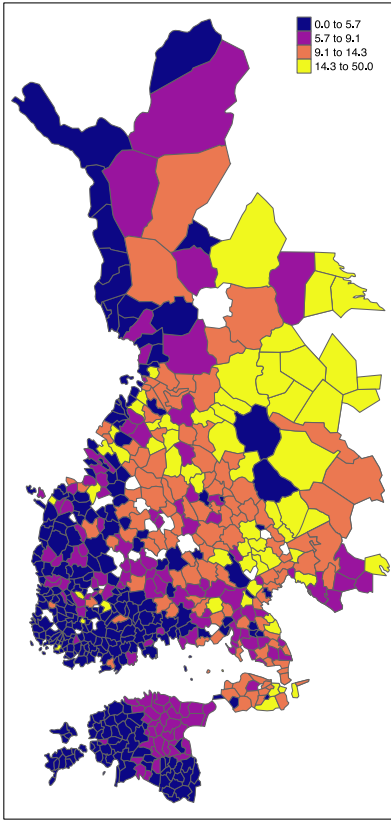


Figure 6: Total percentage of the most common line structures typical of quantitative runosong metre (233, 323, 332) in the runosong texts for each parish in the corpus

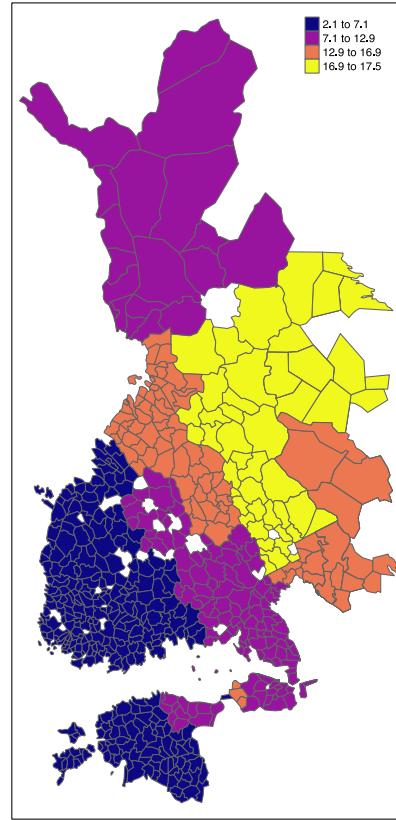


Figure 7: Total percentage of the most common line structures typical of quantitative runosong metre (233, 323, 332) in the runosong texts for each region (county or dialect area) in the corpus

area largely overlaps with the violet area on [Figure 4](#), while the dark blue shading corresponds with the area more associated with accentual metre in Sarv's study. On this basis, we may propose a working hypothesis: in one quarter of the Finnic runosong area (the dark blue areas), accentual metre prevails, while in the other three quarters of this space, features typical of quantitative metre predominate.

These assumptions and hypotheses call for verification through more detailed analysis of the actual metric qualities of runosongs at the level of parish, region, genre, singer, collector and related variables.



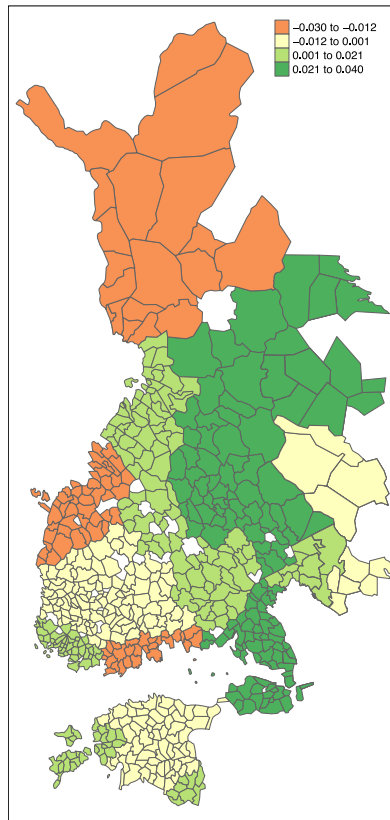


Figure 8: Deviations from expected values (based on a linear regression line) for the percentage of lines of typical quantitative runosong metre with the given average word length

## 6 Conclusion

An immense amount of work will be needed to complete a careful and sound metric analysis of thousands of texts of oral poetry that have been transcribed in different but related languages and dialects, in different orthographies and with different degrees of accuracy. Even if we use computing tools for the metrical analysis, collecting and considering background information about linguistic and folkloric variation, performance styles and the like remain enormous tasks. However, the current study has shown that a large-scale approximation performed by computational analysis can deliver meaningful results in more general terms if we agree to overlook the details.

Finnic runosong metre is generally thought to have originated from a model of quantitative metre (so-called *Kalevala* metre) with specific rules about the placement of stressed syllables based on their quantity. Along with linguistic changes, especially the shortening of words, and under various cultural influences in many regions in the area, the metre gradually lost its quantitative

features and shifted towards an accentual metre in which quantity rules were loosened or disappeared altogether. There is no previous data-based research of this metrical variation across this entire Finnic runosong area. The current study has estimated the extent and geographic nature of variation in the Finnic runosong tradition on the quantitative-accentual scale based on an approximate syllabification of a corpus of approximately 200,000 song texts. To this end, we have used two different indicators: (1) the average word length in syllables by parish/region, and (2) the percentage of the most common line structures typical of quantitative runosong metre by parish/region. The results of both surveys appear to be generally coherent, thus providing the very first large-scale, data-based overview of metric variation in the Finnic runosong tradition.

## Acknowledgments

The data and code used in the current study are available at the Zenodo repository (Sarv et al. 2021).

## References

- Anderson, Walter (1935). *Studien zur Wortsilbenstatistik der älteren estnischen Volkslieder. Eesti Rahvaluule Arhiivi Toimetused 2./Acta et Commentationes Universitatis Tartuensis B XXXIV1*. Tartu.
- Anttonen, Pertti (2005). *Tradition through Modernity. Postmodernism and the Nation-State in Folklore Scholarship*. Helsinki: Finnish Literature Society. DOI: [10.21435/sff.15](https://doi.org/10.21435/sff.15).
- Frog (2019). “The Finnic Tetrameter – A Creolization of Poetic Form?” In: *Studia Metrica et Poetica* 6.1, pp. 20–78. DOI: [10.12697/smp.2019.6.1.02](https://doi.org/10.12697/smp.2019.6.1.02).
- Genetz, Arvid (1884 [1881]). *Suomen kielioppi. Äänne-, Muoto- ja Runous-oppi. Oppikouluja varten. Toinen painos*. Helsinki: K. E. Holm’in kustantama.
- Grünthal, Riho (forthcoming). “The Finnic Languages and Main Dialects”. Revised version from Grünthal & Sarhima 2004 Itämerensuomalaiset kielet ja niiden päämurteet. URL: <https://www.sgr.fi/muutjulkaisut/ItämerensuomalaisetKieletMurteet2012.pdf>.
- Harvilahti, Lauri (2013). “The SKVR Database of Ancient Poems of the Finnish People in Kalevala Meter and the Semantic Kalevala”. In: *Oral Tradition* 28.2, pp. 223–232.
- Helinski, Eugen (1998). *Samojedit ja šamanismi: Viisi luentoa samojedeista, šamanismista ja uralilaisesta kulttuurista*. Ed. by Timo Leisiö Larisa; Leisiö. Tampere: Tampereen Yliopisto, Kansanperinteen Laitos.
- Huttu-Hiltunen, Pekka (2008). *Länsivienalainen runolaulu 1900-luvulla. Kuuden runolaulajan laulutyylin kulttuurisensitiivinen musiikkianalyysi*. Kuhmo: Juminkeko.
- Jakobson, Roman (1979). “Metrics”. In: *Selected Writings V. On Verse, Its Masters and Explorers*. The Hague, Paris, New York, pp. 147–159.

- Kallio, Kati (2021). "Performance, music and meter in Finnic Kalevala-metric poetry". In: *Versification: Metrics in Practice*. Ed. by Frog, Satu Grünthal, Jarkko Niemi, and Kati Kallio. Helsinki: SKS, pp. 59–78. DOI: [10.21435/sflit.12](https://doi.org/10.21435/sflit.12).
- Kallio, Kati, Frog, and Mari Sarv (2017). "What to Call the Poetic Form: Kalevala-Meter, Regivärss, Runosong, Alliterative Finnic Tetrameter, or Something Else?" In: *RMN Newsletter* 12–13, pp. 139–161.
- Klemettinen, Pasi (2006). "Ei se synny synnyttämättä" *Selvitys digitoitiprojektin vaiheista ja työprosesseista*. Suomalaisen Kirjallisuuden Seura. URL: [https://www.finlit.fi/sites/default/files/mediafiles/tutkimus/elias\\_loppuraportti.pdf](https://www.finlit.fi/sites/default/files/mediafiles/tutkimus/elias_loppuraportti.pdf).
- Korhonen, Mikko (1994). "The Early History of the Kalevala Metre". In: *Songs Beyond the Kalevala. Transformations of Oral Poetry. Studia Fennica Folkloristica* 2. Ed. by Anna-Leena Siikala and Sinikka Vakimo. Helsinki: SKS, pp. 75–87.
- Krikmann, Arvo (1997). *Sissevaateid folkloori lühivormidesse I*. Tartu: Tartu Ülikooli kirjastus. URL: <http://haldjas.folklore.ee/~kriku/LEX/KATUS.HTM>.
- Kuusi, Matti (1983). *Maria Luukan laulut ja loitsut*. Suomalaisen Kirjallisuuden Seuran toimituksia 379. Mikkeli.
- Kuusi, Matti (1994). "Questions of the Kalevala Metre". In: *Songs Beyond the Kalevala. Transformations of Oral Poetry. Studia Fennica Folkloristica* 2. Ed. by Anna-Leena Siikala and Sinikka Vakimo. Helsinki: SKS, pp. 41–55.
- Kuusi, Matti and Ülo Tedre (1979). "Regivärssilise ja kalevalamöödulise laulu-traditsiooni vahekorra. Dialog üle lahe". In: *Keel ja Kirjandus* 2, pp. 70–79.
- Laitinen, Heikki (2006). "Runolaulu". In: *Suomen musiikin historia 8: Kansanmusiikki*. Ed. by Anneli Asplund et al. Helsinki: WSOY, pp. 14–79.
- Lauerma, Petri (2001). "Larin Parasken metriikasta". In: *Virittäjä* 1, pp. 44–58.
- Lauerma, Petri (2004). *Larin Parasken epiikan kielellisestä variaatiosta*. Helsinki: SKS.
- Leino, Pentti (1986). *Language and Metre: Metrics and the Metrical System of Finnish. Studia Fennica* 31. Helsinki: SKS.
- Leino, Pentti (1994). "The Kalevala Metre and its Development". In: *Songs beyond the Kalevala. Transformations of Oral Poetry. Studia Fennica Folkloristica* 2. Ed. by Anna-Leena Siikala and Sinikka Vakimo. Helsinki: SKS, pp. 56–74.
- Leino, Pentti (2002 [1975]). "Äidinkieli ja vieras kieli: rahvaanrunouden metriikka". In: *Mittoja, muotoja, merkityksiä*. Helsinki: SKS, pp. 207–230.
- Lord, Albert B. (1960). *The Singer of Tales*. Repr. 2000. Cambridge: Harvard University Press.
- Lotman, Mihhail (1998). "Värsisüsteemidest. Peamiselt eesti ja vene värssi näitel". In: *Akadeemia* 9–10, pp. 1846–1874, 2058–2078.
- Oras, Janika (2008). *Viie 20. sajandi naise regilaulumaailm. Arhiivitekstid, kogemused ja mälestused. Eesti Rahvaluule Arhiivi toimetused* 27. Tartu: Eesti Kirjandusmuuseumi Teaduskirjastus.
- Oras, Janika (2010). "Musical manifestations of textual patterning in Estonian regilaul". In: *Journal of Ethnology and Folkloristics* 4.2, pp. 55–68.
- Oras, Janika (2015). *Vadja ja isuri rahvalaulud*. Tartu: MTÜ Rahvalauluselts Helleo, Estonian Folklore Archives of the Estonian Literary Museum. URL: <http://www.folklore.ee/pubte/eraamat/vadjaisuri/en>.

- Oras, Janika, Liina Saarlo, and Mari Sarv, eds. (2003–2021). *Eesti regilaulude andmebaas*. Tartu: Eesti Kirjandusmuuseumi Eesti Rahvaluule Arhiiv. DOI: [10.15155/9-00-0000-0000-0000-0008FL](https://doi.org/10.15155/9-00-0000-0000-0000-0008FL). URL: <http://www.folklore.ee/regilaul/>.
- Põldmäe, Jaak (1978). *Eesti värsiõpetus*. Tallinn: Eesti Raamat.
- Porthan, Henrik Gabriel (1983). *Suomalaisesta runoudesta 1766–1778 [De poësi Fennica]*. Käänt. Iiro Kajanto. Helsinki: SKS.
- Ross, Jaan and Ilse Lehiste (2001). *The Temporal Structure of Estonian Runic Songs. Phonology and Phonetics 1*. Ed. by Aditi Lahiri. Berlin, New York: Mouton de Gruyter.
- Rüütel, Ingrid (1998). “Estonian Folk Music Layers in The Context of Ethnic Relations”. In: *Folklore* 6. URL: <http://haldjas.folklore.ee/folklore/vol6/ruutel.htm>.
- Saag, Lehti et al. (2019). “The Arrival of Siberian Ancestry Connecting the Eastern Baltic to Uralic Speakers further East”. In: *Current Biology* 29.10, 1701–1711.e16. DOI: <https://doi.org/10.1016/j.cub.2019.04.026>.
- Saarinen, Jukka (2018). *Runolaulun poetiikka. Säe, syntaksi ja parallelismi Arhippa. Perttusen runoissa. Akateeminen väitöskirja*. Helsinki: Helsingin yliopisto. URL: <http://urn.fi/URN:ISBN:ISBN%20978-951-51-3919-1>.
- Sadeniemi, Matti (1951). *Die Metrik des Kalevala-Verses. FF Communications 139*. Helsinki: Suomalainen Tiedeakatemia.
- Sarajas, Annamari (1956). *Suomen kansanrunouden tuntemus 1500–1700-lukujen kirjallisuudessa*. Porvoo: WSOY.
- Sarv, Mari (2008). *Loomiseks loodud: regivärsimõõt traditsiooniprotsessis*. Tartu: EKM Teaduskirjastus.
- Sarv, Mari (2011). “Possible foreign influences on the Estonian regilaul metre: language or culture?” In: *Frontiers in Comparative Prosody*. Ed. by Mihhail Lotman and Maria-Kristiina Lotman. Linguistic Insights; 113. Peter Lang Verlag, pp. 207–226.
- Sarv, Mari (2015). “Regional Variation in Folkloric Meter: The Case of Estonian Runosong”. In: *RMN Newsletter* 9, pp. 6–17.
- Sarv, Mari (2019). “Poetic metre as a function of language: linguistic grounds for metrical variation in Estonian runosongs”. In: *Studia Metrica et Poetica* 6.2, pp. 102–148. DOI: [10.12697/smp.2019.6.2.04](https://doi.org/10.12697/smp.2019.6.2.04).
- Sarv, Mari (2020). “From tradition to data: The case of Estonian runosong”. In: *Arv. Nordic Yearbook of Folklore* 76, pp. 105–117.
- Sarv, Mari, Kati Kallio, Maciej Janicki, and Eetu Mäkelä (2021). *Metrical variation in the Finnic runosong tradition: challenges and possibilities for a computational approach (code + dataset) (Version 1.0)*. [Dataset]. Zenodo. DOI: [10.5281/zenodo.4537606](https://doi.org/10.5281/zenodo.4537606).
- SKVR (2021). *SKVR-tietokanta – kalevalaisten runojen tietokanta*. Suomalaisen Kirjallisuuden Seura. URL: <http://skvr.fi> (visited on 07/14/2021).
- Tampere, Herbert, Erna Tampere, and Otilie Kõiva (2016). “Eesti rahvamuusika antoloogia. Veebiväljaanne”. In: *Helisalvestusi Eesti Rahvaluule Arhiivist* 3. Ed. by Janika Oras and Kadi Sarv. Tartu: Eesti Kirjandusmuuseumi Teaduskirjastus. URL: <http://www.folklore.ee/pubte/eraamat/rahvamuusika/en/>.

- Tampere, Herbert, Erna Tampere, and Otilie Kõiva (2017). “The Field of Song and the Four-Legged Horse: On the Dialogue of Genres in Kalevala-Meter Poetry”. In: *Classics@: Singers and Tales in the 21st Century; The Legacies of Milman Parry and Albert Lord*. Harvard: Center for Hellenic Studies, Harvard University. URL: <https://chs.harvard.edu/CHS/article/display/6617>.
- Tarkka, Lotte, Eila Stepanova, and Heidi Haapoja-Mäkelä (2018). “The Kalevala’s Languages: Receptions, Myths, and Ideologies”. In: *Journal of Finnish Studies* 21.1–2, pp. 15–45. URL: <http://hdl.handle.net/10138/301432>.
- Viidalepp, Richard (1959). “Eesti rahvalaulude poeetika ja keel”. In: *Eesti rahvaluule ülevaade*. Ed. by Richard Viidalepp. Eesti Riiklik Kirjastus: Tallinn, pp. 116–177.