# The SKVR Database of Ancient Poems of the Finnish People in Kalevala Meter and the Semantic Kalevala

#### Lauri Harvilahti

## **Introduction: The Folklore Archives of the Finnish Literature Society**

The Folklore Archives are the central holdings for oral tradition research in Finland. The actual collection of folklore began in Finland in the first half of the nineteenth century with poems and charms in the Kalevala meter. Since the mid-nineteenth century, the press also participated by publishing appeals to collectors, and these calls were invariably met with great enthusiasm. Towards the end of the nineteenth century a network of collectors was established, guidelines for collection were prepared, and folklore collection was encouraged in general. At the turn of the century, the Kalevala-meter poems and charms were joined by collections of folktales, in the 1930s legends were added, and gradually all fields of agrarian folklore, proverbs and riddles, the belief tradition, and laments were included. In 1900 the archive of the Finnish Literature Society still had only approximately 200,000 items of folklore, and today the archive holds approximately four million folklore items. In 1937 the Society's folklore collections were consolidated into a research institution known as the Folklore Archives of the Finnish Literature Society. Various organizations and educational establishments also responded and took an active interest in folklore collecting.

The Folklore Archives seek to collect oral tradition, personal narratives, and memory lore in a number of different ways. Nowadays this work is primarily focused on organizing collection campaigns and fieldwork across the country. The archives also actively maintain their own respondent network.<sup>2</sup> Around 30,000 people have been involved in collecting this material. Geographically the collection work focuses primarily on the cultural area of Finland and Karelia (<a href="http://www.finlit.fi/english/kra/">http://www.finlit.fi/english/kra/</a>). The collections encompass oral tradition, folk music, ethnological descriptions, and oral history/memory lore.

<sup>&</sup>lt;sup>1</sup> See Harvilahti 2003 and 2012.

<sup>&</sup>lt;sup>2</sup> See Vento 1998.

# The SKVR Corpus of Kalevala-Metric Poetry (<a href="http://dbgw.finlit.fi/skvr/">http://dbgw.finlit.fi/skvr/</a>)

The poems in the Kalevala-meter were collected largely thanks to the initiative of the Finnish Literature Society (established already in 1831). There are now altogether some two million lines of Kalevala-metric poetry in the archives of the Finnish Literature Society, collected primarily in the nineteenth century and the first half of the twentieth century. The typical poetic devices of Kalevala-metric poetry are as follows: the use of alliteration and assonance, the verse structure of eight syllables, the trochaic meter, and the rules of syntactic parallelism. Naturally the performers were not aware of the finer distinctions, but they did observe the basic register of Kalevala poetry: together these primary features formed a poetic culture observing a fairly uniform poetic system. This system was shared by Finnish, Karelian, and Estonian performers of oral poetry in Kalevala meter.

The Society's first secretary, Elias Lönnrot, compiled the Finnish national epic Kalevala on the basis of such collected poetry. The first edition of the *Kalevala* appeared in 1835, and the second and greatly enlarged edition in 1849 (see further Harvilahti 2008). An edition of documented poetic texts, Suomen Kansan Vanhat Runot (SKVR), was later produced in 34 volumes in 1908-48 and 1997. This edition presents approximately 89,000 poetic texts requiring almost 27,000 pages. The Finnish Literature Society began digitalizing SKVR in 1998. The conversion of the texts into digital format by scanning and OCR (Optical Character Recognition) was assigned to an Estonian team, as Estonia already had experience with such work. All printed volumes were digitized with character-by-character equivalence to SKVR, appearing as an early major open-access corpus in XML format.<sup>3</sup> The greatest advantage of digital text is naturally that it permits efficient, comprehensive searches of an entire corpus for which a database is required. XML also permits the easy transfer of data to many different applications. In its present form, the corpus has a rather basic user interface allowing searches of the texts by keywords (largely according to grammatical inflection and the sometimes inconsistent orthographic forms used in transcribing performed texts), and by such metadata as collectors, geographical names, and dates. Currently, the database does not fully support multidisciplinary applications, exists only in Finnish, and is somewhat tedious to use. In the very near future a standard thematic index will be added to the corpus. Every poem will be anchored to this index, thereby greatly increasing the search possibilities for use within linguistic and poetic analyses.

The new database that will be realized in the near future shall:

- Contain the above-mentioned standard thematic index
- Enhance the capability for making searches
- Enable the use of a number of parameters that have not been included in the present version

Furthermore, the indexes may also be attached to narrative instances (or macro-propositions) as metadata. In this way an index might be produced that contains lists of objects that share

<sup>&</sup>lt;sup>3</sup> See Saarinen 2001.

common attributes of semantic metadata, as has already been done in the so-called Semantic Kalevala project.

### The Database of Estonian Oral Poetry (<a href="http://www.folklore.ee/regilaul/avaleht/">http://www.folklore.ee/regilaul/avaleht/</a>)

The development process of the SKVR corpus has given rise to lively international cooperation. One of the fruits of this collaboration is the corpus of old Estonian Kalevala-metric poetry, *Eesti Regilaulude Andmebaas* ("Estonian Runic Songs Database"), launched in 2010. This Estonian database can thus work together with its Finnish counterpart to form a very handy tool for making comparisons between the Kalevala traditions of Finland and Estonia. They are both relatively easy to use, provided that the user knows the languages, including the dialects and the poetic register. One of the central long-term aims is to build an interface that will allow searches across both Estonian and Finnish digital corpora of Kalevala-metric poetry.

#### The Semantic Kalevala and the SKVR

The Semantic Kalevala has been created by the Semantic Computing Research Group of the Helsinki University of Technology (Aalto University) and the University of Helsinki as part of the larger semantic web portal called Culture Sampo (<a href="http://www.kulttuurisampo.fi">http://www.kulttuurisampo.fi</a>). The work was part of the national web ontology project FinnONTO, initiated and headed by professor Eero Hyvönen. The general goal of this large project was to combine the benefits and synergy of Web 2.0 and semantic web technologies, and to demonstrate the results in various semantic web portals and applications. The semantic web forms a service that enables the creation of a content-based metadata level for making searches. In this way, the semantic web complements and improves the capabilities of conventional use of the web.

The first phase was an annotation of the *Kalevala* (http://www.seco.tkk.fi/applications/kulttuurisampo/kalevala/). During the annotation process the epic was divided into three levels of narrative hierarchy: events, scenes, and episodes. Each event belongs to a scene, whereas scenes belong to episodes that are subplots within the epic narrative itself. By browsing these different levels, the user can follow the plot of the epic. In addition, the Semantic Kalevala includes descriptions of the epic's main characters and places, as well as concise summaries (or macropropositions) of the scenes. These additions help a generalist user to understand the archaic expression of the Kalevala-metric poetry and to follow the elaborate narration of the epic, compiled in its literary form by Elias Lönnrot.

In the very near future the Semantic Kalevala will be published as a SPARQL endpoint service in the Linked Open Data publishing platform of the Linked Data Finland project. This project is a successor to the FinnONTO project. The SPARQL endpoint is provided for querying and managing data, and it constitutes the basis for application development. The first version is available at <a href="http://www.ldf.fi/dataset/folklore/index.html">http://www.ldf.fi/dataset/folklore/index.html</a>.

The next phase of this process would be to annotate both editions of the *Kalevala* produced by Elias Lönnrot (1835, 1849) and part of the SKVR corpus more accurately according to the principles of the primary features of Kalevala-metric poetry. In other words, there exists a

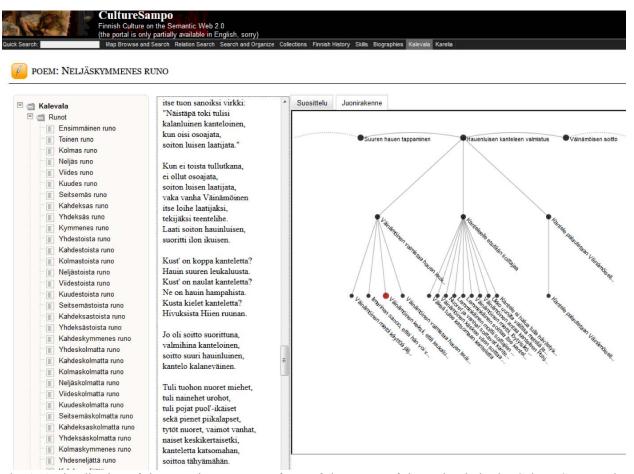


Fig. 1. A visualization of the narrative structure of one of the poems of the *Kalevala* in the Culture Sampo. The cantos (poems) of the *Kalevala* are listed on the left, the text of the beginning of the fortieth canto is in the middle, and the plot structure of the episodes of that canto on the right, in this case the origin of the *kantele* (a musical instrument) made of fish bones. Image reproduced (with permissions) from <a href="http://www.kulttuurisampo.fi/kulsa/kalevala.shtml?itemUri=http%3A%2F%2Fkulttuurisampo.fi/kulsa/kalevala.shtml?itemUri=http%3A%2F%2Fkulttuurisampo.fi/kulsa/s23Instance\_ID1220963995343&fromplot=true&selectedScene=http%3A%2F%2Fkulttuurisampo.fi/w2Fannotaatio%23Instance\_ID1221041435134#221-226.

need to develop an annotation scheme that can take into account the shared common attributes of semantic metadata of the Kalevala-metric poetry. One aim of combining the semantic web methodology and the textual analysis of the poetic language is to trace the editorial selections of Elias Lönnrot. A long-term goal is to produce a *Critical eEdition of the Kalevala* by carefully analyzing the *Kalevala* and the Finnic traditional oral versions of the Kalevala-metric poetry that Lönnrot most probably had at his disposal.

# The Origin of Kantele/Kannel, a Brief Case Study

The *kantele*, or *kannel*, is the most widespread ethnic instrument among the Finnic peoples. A number of scholars believe that the mythological epic poem on the *Origin of the Kantele* belongs to the oldest stratum of Finnic mythological epics, and according to Anna-Leena Siikala (2012:365-66, 449-51), this poem is part of the common Balto-Finnic mythological heritage that can be traced to earlier "mythological dialects." Using the terminology of Lauri

Honko, the Finnish and Estonian epics form a kind of ecosystem of milieu-morphological adaptation.<sup>4</sup> According to John Miles Foley the traditional registers function in oral texts as storage containers for idiomatic means of communication.<sup>5</sup> The Balto-Finnic mythology is a multilayer tradition and its mythological registers contain elements of an old Eurasian worldview and subsequent development of models of cultural thinking (Siikala 2012:475).

In the poems collected from northern areas of Finnish and Russian Karelia the *kantele* made of fish bones is prevalent; this version is known among the Finnic ethnic groups around St. Petersburg as well. In southern parts of the Gulf of Finland among the so-called Ingrian Finns the most common version of the poem is the *kantele* made of horns from either an elk or ram. In Estonian material there are some surprisingly archaic versions of the *kantele* poems that refer to the common mythological epic poems of the Finnic ethnic groups.

The *kantele* of the poems has also been made of a number of other materials. Among the materials that have been mentioned as being used for the lid and the bottom of the instrument are trees—birch, spruce, oak, and maple—and also steel. In Estonia the song has mostly been preserved in a more lyric-epic form, whereas in Finland the song is clearly a mythological epic. There are hundreds of parallels in the two databases. In order to avoid lengthy comparisons, I will demonstrate the use of the Finnish and Estonian databases (<a href="http://dbgw.finlit.fi/skvr/">http://dbgw.finlit.fi/skvr/</a> and <a href="http://www.folklore.ee/regilaul/avaleht/">http://www.folklore.ee/regilaul/avaleht/</a>) by taking as an example poems that contain one particular formula—*The kantele/kannel got finished*—in both traditions.

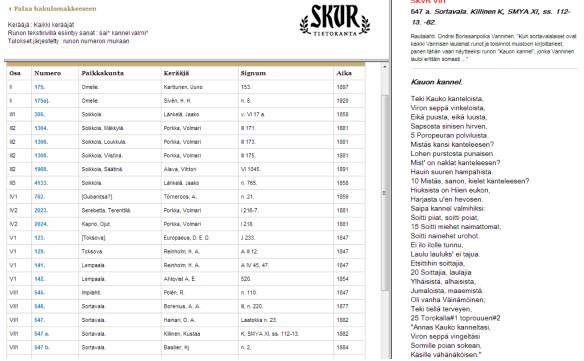


Fig. 2. A screenshot from the SKVR database: a search of poems containing the formula *The kantele got finished*. (The list of poems appears on the left, and the text of our example is on the right.)

<sup>&</sup>lt;sup>4</sup> See Honko 1998:101-02 et passim.

<sup>&</sup>lt;sup>5</sup> See, for instance, Foley 1995:49 et passim, 2002.

First, we have an instance of the *Origin of the Kantele* poem, collected in 1882 from the Eastern part of Karelia (Ladoga Karelia)—nowadays belonging to the Russian Federation—by K. Killinen from the singer Ontrei Vanninen (SKVR VII1, 547A: verses 1-13, 17-29):<sup>6</sup>

Teki Kauko kanteloista,
Viron seppä vinkeloista,
Eikä puusta, eikä luusta,
Sapsosta sinisen hirven,
Poropeuran polviluista.
Mistäs kansi kanteleesen?
Lohen purstosta punaisen.
Mist' on naklat kanteleesen?
Hauin suuren hampahista.
Mistäs, sanon, kielet kanteleesen?
Hiuksista on Hiien eukon,

Saipa kannel valmihiksi:

Harjasta u'en hevosen.

. . .

Ei ilo ilolle tunnu, Etsittihin soittajia, Oli vanha Väinämöinen; Teki tiellä terveyen, Torokalla toprouuen "Annas Kauko kanneltasi, Viron seppä vingeltäsi Sormille poian sokean, Käsille vähänäköisen." Kauko shaped a kantele Estonia's smith a curved thing

Neither of wood nor of bone
But of a blue elk's shoulder
A reindeer's knee-bones
Whence the kantele's lid?
Made from a red salmon's tail
And whence the kantele's pegs?
From the teeth of a great pike

And whence, I say, the kantele's strings?

the hair of a demon's dame From the mane of a stud-horse

The kannel got finished

. . .

Joy had not the feel of joy—
There was a search for singers—

It was old Väinämöinen Gave a greeting on the road, Said good morning on the lane: "Give, Kauko, your kantele, Estonia's smith your curved thing

into the blind boys fingers,

To the dim-sighted one's hands."

In Estonian parallels of the *Origin of the Kantele/Kannel* several poems contain almost identical episodes. Sometimes even the formulas are almost the same, in some cases word-forword equivalent verses can be found. There is no lack of similar examples. A detailed analysis of the digital corpora would bring more evidence of a long-term, close relationship between the Estonian and Finnish oral poetry. This relationship is remarkable since there is, after all, an essential difference between Estonian and Finnish, although the languages belong to the same Balto-Finnic group. I cite a representative chapter telling about making the parts of the *kannel*. The poem was collected from Pärnu, in southwestern Estonia, relatively far from Finland (Eesti regilaulude andmebaas H II 43, 903/4 (18): verses 4-17):<sup>7</sup>

<sup>&</sup>lt;sup>6</sup> The English translation here is adapted from that of Keith Bosley found in the edition by Kuusi (1995:71).

<sup>&</sup>lt;sup>7</sup> The English translation is my own.



Fig. 3. A search for the formula *The Kannel got finished* from the Estonian Runic Songs Database. The list of results appears on the left, and the chosen example is on the right.

Kellest teen ma kannelit?
Vihterista, vahterista.
Kellest teen ma kandle põhja?
Õunapuusta õigeesta.
Kellest teen ma kandle kaane?
Sarapuusta sirgeesta.
Kellest teen ma kandle vitsad?
Need teen lõhe lõua luusta.
Kellest teen ma kandle kaela?
Selle teen põdra põlve luusta.
Kellest teen ma kandle pulgad?
Need teen havi hammastesta.
Kellest teen ma kandle keeled?
Juukseesta sirgeesta.
Sai see kannel valmissa

Otsisin ma mängijada. . . .

From what will I make the kannel?
From a twig, from a maple.
What the bottom of the kannel?
From an upright apple tree.
What the lid of the kannel?
From a straight walnut tree.
What the hoops of the kannel?
Those from the jaw-bones of salmon.
What the neck of the kannel?
From the kneebone of a deer.
What the pegs of the kannel?
Those from teeth of a pike.
What the strings of the kannel?
From a straight hair.
The kannel got finished

It is really interesting to find so many straight formulas (Lord 1960:30, 35, 46) in the Finnish and Estonian examples. In some cases there is a difference between the mythical content of the Finnish poems and the lyric-epic tone of the Estonian parallel. For instance, the strings of the instrument are made in the Estonian poem simply from straight hair, but in Karelia from hair

I looked for a player. . . .

of a demon's dame. However, in both instances the word-for-word equivalent statement-type formula *The kantele/kannel got finished* concludes the episode of preparing the parts of the instrument.

Oral poetry forms a culture-bound model of communication, a large semantic network suited to the expression of different meanings. Oral performers and compilers of epics alike draw on their store of formulas within the confines of the traditional rules and their personal competence. The study of the register of Estonian and Finnish epics serves as a laboratory for testing culture-bounded models of communication in the shared heritage of the Kalevala-metric system. The real problem is that at the moment the Finnish and Estonian databases do not have translations for generalists and the sites are usable primarily by specialists with good command of not only the language, but also the poetic register.

One way to combine the modern web technology and the traditional folkloristic methods would be to use international motif-indexes, in which the basic idea underlying folk products regarded as belonging together (epics, folktales, legends, and so forth) is expressed concisely: a verbal definition is supplemented by letter/number codes for each type. The origin of mythical music instruments is an internationally known narrative motif. In the Motif-Index of Folk-Literature of Stith Thompson it figures as: A1461.2 Origin of Lyre (1955:232; see also D1441.1.3.2 Magic musical instrument calls animals together [1956:231]). The advantage of the type-indexes is (in principle) their universality: numerical codes are independent of language. In practice, numerous codes are needed simultaneously for classifying the majority of folklore variants, and new types to which no existing code applies could always be added. The potential of the digital corpora, the semantic web technologies, and web-based type-indexes might enable the creation of a challenging model for research on oral poetry.

### **Summary**

At the moment the open-access databases of Estonian and Finnish oral poetry (http:// dbgw.finlit.fi/skvr/ and http://www.folklore.ee/regilaul/avaleht/) are primarily usable for specialists who are familiar with the languages involved. There is no English search possibility for the Finnish SKVR corpus or its Estonian equivalent, and it is very difficult to navigate the pages as a non-specialist. I would suggest that in the future at least the existing Estonian and Finnish databases of Kalevala-metric folk poetry could be linked using, for example, the approach of semantic web ontology. This approach could open the sites for generalists as well and would be made possible by using the potential of the metadata schemes that are machinereadable—and language independent. The new type of platform could be applicable for a fully multidisciplinary range of studies in linguistics, folkloristics, literary studies, history, and various interdisciplinary textual studies. The development and application of the common scheme could be undertaken in conjunction with comparative textual research on oral traditions in order to test and develop the scheme's relevance and applicability for diverse types of research priorities. There is a need for a versatile www-infrastructure that would provide a platform for epic corpora (and broader corpora of oral traditions) with an ergonomic user-interface allowing searches within an epic or cultural epic tradition, and across the epic and poetic traditions of multiple

cultures according to a diverse range of criteria. Citing a discussion between Eero Hyvönen and myself: based on such repositories as the Semantic Kalevala and data services, distributed possibly in different organizations in different countries, the idea of linked epics and oral tradition could be put forward in a new way. The result could be a step toward a broader understanding of shared features exhibiting diverse contextual meanings in epic traditions elsewhere in Eurasia.

The Finnish Literature Society, Helsinki

#### References

Foley 1995	John Miles Foley. <i>The Singer of Tales in Performance</i> . Bloomington: Indiana University Press.
Foley 2002	How to Read an Oral Poem. Urbana: University of Illinois Press.
Harvilahti 2003	Lauri Harvilahti. "Folklore and Oral Tradition." <i>Oral Tradition</i> , 18.2:200-02. <a href="http://journal.oraltradition.org/issues/18ii/harvilahti">http://journal.oraltradition.org/issues/18ii/harvilahti</a>
Harvilahti 2008	"Kalevala." In <i>The Greenwood Encyclopedia of Folktales and Fairy Tales. Volume 2: G-P.</i> Ed. by Donald Haase. Westport, CT: Greenwood Press. pp. 529-30.
Harvilahti 2012	"Finland." In <i>A Companion to Folklore</i> . Ed by Regina F. Bendix and Galit Hasan-Rokem. Hoboken, NJ: Wiley-Blackwell. pp. 391-408.
Honko 1998	Lauri Honko. <i>Textualising the Siri Epic</i> . Folklore Fellows Communications, 264. Helsinki: Academia Scientiarum Fennica.
Kuusi 1995	Matti Kuusi, ed. <i>A Trail for Singers: Finnish Folk Poetry—Epic.</i> Helsinki: Finnish Literature Society.
Lord 1960	Albert B. Lord. <i>The Singer of Tales</i> . Cambridge, MA: Harvard University Press.
Saarinen 2001	Jukka Saarinen. "Kalevalaic Poetry as a Digital Corpus." FF Network for the Folklore Fellows, 22:6-9.
Siikala 2012	Anna-Leena Siikala. <i>Itämerensuomalaisten mytologia</i> . Helsinki: Suomalaisen Kirjallisuuden Seura.
Thompson 1955	Stith Thompson. Motif-Index of Folk-Literature. A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla,

	Fabliaux, Jest-Books and Local Legends. Volume I: A-C. Rev. and enlarged ed. Copenhagen: Rosenkilde and Bagger.
Thompson 1956	Motif-Index of Folk-Literature. A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-Books and Local Legends. Volume II: D-E. Rev. and enlarged ed. Copenhagen: Rosenkilde and Bagger.
Vento 1998	Urpo Vento. "The Folklore Activities of the Finnish Literature Society." <i>FF Network for The Folklore Fellows</i> , 15:5-7.