

# Mieleni minun tekevi vai tekevikö sittenkään eli uudenlaisen kalevalamitan rakentelua

## Taustaa

Kalevala on Suomen kansalliseepos. Kaikki suomalaiset ovat sen lukeneet jo montakin kertaa kannesta kanteen. Paitsi minä.

Tämä asia on tuottanut itselleni suurta häpeää jo vuosikymmenien ajan. En häpeä Kalevalaa enkä sitä, että muut ovat sen lukeneet vaan sitä, etten itse ole sitä lukenut. Pidän itseäni moukkana. En kehtaa kohdata lasteni, naapureideni tai työkavereideni syyttäviä katseita. He tietävät sen. He näkevät sen alistuneesta, arkana pälyilevästä olemuksestani ja laahaavasta, epävarmasta askelluksestani.

Noin kymmenen vuotta sitten seisoin antikvaarisessa kirjakaupassa erään kirjahyllyn edessä. Siellä se kökötti ja katsoi suoraan kohti. Vuonna 1958 painettu, Akseli Gallen-Kallelan kuvituksilla varustettu upea Kalevala. Otin kirjan käteeni ja avasin ensimmäisen lehden näkyviin. Aukeamalla luki: *"Paikkarin Eliaksen muistolle."*

Tämä kosketti syvältä. Tekikö Paikkarin torpan kasvatti Elias Lönnrot 1800-luvulla siis runonkeruumatkinsa turhaan? Kiersikö hän syyttä suotta Hämeessä, Savossa, Karjalassa, Kuolassa ja Virossa? Kohtasiko hän tarpeettomasti matkoillaan vastuksina koleraa, lavantautia, isorokkoa ja tulirokkoa? Puhumattakaan tuskallisista miesflunssista ja kynsivallintulehduksista, joista historiankirjatkin vaikenevat. Paljain jaloin taivalsi reissuillaan Elias. Kuinka tämän kaiken kestävätkään hänen kantapäätänsä, sitä voi vain arvailla sivellessäni pumppupullosta rasvaa omiin kuivettuneisiin kantapäihini?

*"Mikä tässä maksavi?",* aivoni ajattelevi ja puhe'eni putoelevi. *"Enköpä minä mahtimieli, kansasta Kalevan kajauttasi, lajivirren laatijasta laulaisi?",* sanat suussani sulivat, kielelleni kerkisivät ja hampahilleni hajosivat.

Kuulin äänen kuiskaavan korvaani: *"Et velikulta veikkoseni, et ilmoisna ikänä, kuuna kullon valkeana, eeposta Eliaksen ennätä, loitsuja Lönnin lukea! Vempula olet, veltto vainen, laiskuutehen taipuvainen!"*

Silloin suutuin suomalainen, ponteva pohjalainen parahdin! Nyrkkini pöytähän pölistin, hiuksiani harvoja hivelin, rintaani raastaen riistin, kiron kipakan kiljaisin: *"Häpeääni hämmästelen, saamattomuuttani sadattelen! Opuksen oivan ostan, maltaiset mahtavat maksan! Luenpa luvut lukuisat, runot rujot, rukihiset!"*

Ostin siis tuon Kalevalan yli kymmenen vuotta sitten. Siinä on 729 sivua. Olen 16.2.2022 edennyt sivulle 203. Kirjanmerkkinä on maaliskuulta 2013 oleva pieni mainoslehtinen. Ilmeisesti olen siis aloittanut lukuurakkani aika tarkkaan 9 vuotta sitten ja kyennyt sinä aikana reilusti alle kolmasosan kirjasta etenemään. Onko se muka paljon?

Kalevalanpäivä 28.2.2022 lähestyy. Voisin asettaa tavoitteeksi lukea kirja loppuun siihen mennessä. Siihen on tällä hetkellä 12 päivää. Nämä ovat hetkiä, jolloin veri punnitaan. Annan periksi. En pysty. Minusta ei ole. On se Kalevala niin kovin kalevalainen sisällöltään.

Jos Kalevala ei innosta lukemaan, niin olisiko mahdollista löytää teos, joka muistuttaisi mahdollisimman paljon sitä, olematta kumminkaan karmaiseva Kalevala? Kykenisikö humanismiin taipuvainen luonnontieteilijä löytämään kalevalaisuuden syvimmän olemuksen?

Ei tietenkään kykene. Mutta jotakin hänkin voi tehdä.

## Kirjainten esiintymistiheys

Project Gutenberg on verkkosivusto, jossa on yli kuusikymmentätuhatta sähköistä kirjaa. Suuri osa sinne tallennetuista kirjoista on todellisia klassikkoteoksia. Esimerkkejä suomenkielisistä kirjoista ovat mm. Johan Ludvig Runebergin Vänrikki Stoolin tarinat, Aleksis Kiven Seitsemän veljestä, Minna Canthin Työmiehen vaimo ja näiden lisäksi satoja muita. Valinnanvaraa siis on yllin kyllin lukea ja jättää lukematta.

Sivustolle talletetut kirjat on mahdollista avata tekstimuodossa ja näin ollen siirtää esimerkiksi Exceliin. Excelin toimintojen avulla on mahdollista analysoida tilastollisesti esimerkiksi eri kirjainten esiintymistiheydet tekstissä. Lähdin siis tutkimaan eri kirjojen ”kalevalaisuusastetta” eli eräänlaista uudenlaista kalevalamittaa.

Aluksi selvitin Kalevalassa (Lönnrot 1835) esiintyneiden kirjainten määrän ja tämä tulos on esitettyinä taulukossa 1. Yhteensä kirjassa oli 451586 kirjainta. Mielenkiintoista on havaita, että kuusi yleisimmin esiintynyttä kirjainta A, I, N, E, T ja S kattavat yhteensä jo noin 56 % kaikista Kalevalan kirjaimista. Vastaava havainto, tosin hieman eri järjestyksessä, voidaan tehdä yleisemminkin suomenkielisistä teksteistä. Tästä löytyy ainakin yhden asiantuntijan tekemä tutkimuskin (Pääkkönen 1991).

*Taulukko 1. Kirjainten esiintymistiheydet Kalevalassa.*

Kirjain	kpl	%
A	52998	11,74 %
I	51117	11,32 %
N	42527	9,42 %
E	37424	8,29 %
T	36878	8,17 %
S	30982	6,86 %
L	30601	6,78 %
K	27040	5,99 %
O	24835	5,50 %
U	21908	4,85 %
Ä	20335	4,50 %
H	13605	3,01 %
M	13563	3,00 %
V	13325	2,95 %
P	11076	2,45 %
R	10565	2,34 %
J	5655	1,25 %
Y	5011	1,11 %
Ö	1881	0,42 %
G	221	0,05 %
D	39	0,01 %
B	0	0,00 %
C	0	0,00 %
F	0	0,00 %
Q	0	0,00 %
W	0	0,00 %
X	0	0,00 %
Z	0	0,00 %
Å	0	0,00 %
<b>YHTEENSÄ</b>	<b>451586</b>	<b>100,00 %</b>

Tein vastaavan tutkimuksen kaikista alla mainituista kirjoista. Käytän luettelossa esiintyvää numerointia myöhemmissä taulukoissa.

- 1 Kalevala (Lönnrot 1835)
- 2 Kanteletar (Lönnrot 1840)
- 3 Väinölin tarinat (Runeberg 1860)
- 4 Seitsemän veljestä (Kivi 1870)
- 5 Juha (Aho 1911)
- 6 Työmiehen vaimo (Canth 1885)
- 7 Helkavirsiä (Leino 1916)
- 8 Mirdja (Onerva 1908)
- 9 Satuja lapsille luettavaksi (Swan 1901)
- 10 Tulitikkuja lainaamassa (Lassila 1910)
- 11 Punainen viiva (Kianto 1909)
- 12 Putkinotko (Lehtonen 1920)
- 13 Götiska rummen (Strindberg 1904), ruotsinkielinen kirja
- 14 Hemsöborna (Strindberg 1887), ruotsinkielinen kirja
- 15 Wuthering Heights (Brontë 1847), englanninkielinen kirja
- 16 The Adventures of Sherlock Holmes (Conan Doyle 1892), englanninkielinen kirja

Tämän lisäksi mielenkiinnon vuoksi tutkin myös

- 17 Teksti-TV:n sivulla 16.02.2022 esiintyneet uutistekstit osioissa Kotimaa, ulkomaat, talous ja alueuutiset

Laskin Excelin avulla kirjainten esiintymistiheydet. Koska kirjoissa oli eri määrä kirjaimia, skaalasin tulokset vastaamaan 1000 kirjaimesta koostuvaa tekstiä ja tulokset olivat taulukoiden 2 ja 3 mukaiset.

*Taulukko 2. Kirjainten esiintymistiheydet kirjoissa 1 - 9*

Kirjain	1	2	3	4	5	6	7	8	9
A	117,36	121,26	115,74	122,57	115,00	119,62	109,95	112,01	113,44
I	113,19	113,34	104,91	108,88	109,53	104,90	112,44	113,98	117,62
N	94,17	95,47	105,17	91,14	103,33	104,46	91,38	99,60	89,66
E	82,87	77,39	69,80	78,71	83,46	76,23	74,79	79,70	76,54
T	81,66	82,86	83,00	85,63	88,51	89,67	84,62	91,52	88,93
S	68,61	68,32	71,43	71,39	70,83	66,84	64,05	70,30	70,88
L	67,76	69,21	53,80	56,63	56,49	52,11	62,66	58,16	59,33
K	59,88	62,00	58,77	57,84	56,52	56,36	63,34	50,82	62,19
O	55,00	52,76	53,02	45,11	44,45	55,86	50,62	44,79	45,03
U	48,51	52,46	56,42	48,78	45,87	49,17	58,78	50,59	53,85
Ä	45,03	43,20	51,55	54,70	65,47	59,06	42,13	60,01	57,20
H	30,13	25,63	29,89	27,60	28,73	27,94	30,74	24,71	25,18
M	30,03	31,38	31,18	37,95	31,68	33,92	36,49	37,94	27,73
V	29,51	26,40	26,15	23,43	21,32	20,48	25,75	20,56	25,97
P	24,53	23,56	17,86	20,76	16,11	20,51	23,72	13,64	23,17
R	23,40	23,82	24,50	23,14	17,59	20,38	28,46	20,39	21,27
J	12,52	12,84	17,03	21,35	19,59	16,19	12,95	20,42	17,44
YÖGBCFQWXZÅ	15,84	18,10	29,78	24,37	25,52	26,30	27,13	30,86	24,56

Taulukko 3. Kirjainten esiintymistiheydet kirjoissa 10 - 17

Kirjain	10	11	12	13	14	15	16	17
A	117,13	121,17	123,81	94,12	93,39	77,79	81,43	124,17
I	120,13	114,29	102,35	50,19	42,34	71,60	69,85	100,31
N	97,78	86,50	97,89	88,73	90,03	71,77	66,39	86,43
E	73,68	71,15	81,69	97,23	87,67	128,08	122,98	78,01
T	93,81	90,83	88,79	81,74	80,83	86,26	90,34	93,56
S	79,23	71,54	70,82	61,23	65,95	61,16	62,97	71,80
L	55,13	55,82	59,35	47,77	48,86	41,49	39,74	58,65
K	52,79	62,90	58,35	32,01	32,73	7,83	8,21	56,60
O	46,24	43,90	45,15	43,04	49,61	73,66	77,60	58,94
U	46,99	50,45	48,91	19,86	20,33	29,69	30,30	59,46
Ä	58,00	58,84	61,06	23,10	20,44	0,00	0,00	38,09
H	27,58	24,02	24,91	28,16	31,04	65,64	67,32	17,29
M	26,43	27,59	28,44	33,35	34,85	26,51	27,33	29,49
V	22,77	24,71	19,84	25,89	22,74	9,18	10,28	26,40
P	19,63	25,32	21,93	13,76	17,96	15,58	15,76	21,74
R	17,99	26,22	21,30	77,71	71,80	58,70	56,79	25,48
J	19,23	18,51	20,39	9,64	7,96	1,10	1,05	18,90
YÖGDBCFQWXZÅ	25,47	26,25	25,04	172,48	181,50	173,96	171,66	34,67

Taulukoissa olevat luvut voidaan siis tulkita myös promilleiksi eli tuhannesosiksi.

Taulukon alimpana rivinä on harvinaisimpien kirjaimien Y, Ö, G, D, B, C, F, Q, W, X, Z ja Å kooste. Näitä kirjaimia esiintyy Kalevalassa sen verran vähän (jos ollenkaan), että ne on tilastollisen analyysin toimivuuden kannalta syytä yhdistää yhdeksi isommaksi joukoksi.

Kalevalaisuuden mittana voidaan pitää tilastollista tunnuslukua nimeltä  $\chi^2$  (khiin neliö). Tämä lasketaan siten, että kunkin kirjaimen kohdalla lasketaan erotus kyseisen kirjan lukeman ja Kalevalan vastaavan lukeman välillä. Tämä erotus korotetaan toiseen potenssiin ja sen jälkeen jaetaan vielä Kalevalan vastaavalla lukemalla. Tämä toistetaan kaikkien kirjaimien kohdalla ja lasketaan tulokset yhteen.

Niinpä esimerkiksi kirjan 2 eli Kantelettaren "kalevalamitta"  $\chi_2^2$  saadaan laskutoimituksella

$$\chi_2^2 = \frac{(121.26-117.36)^2}{117.36} + \frac{(113.34-113.19)^2}{113.19} + \dots + \frac{(18.10-15.84)^2}{15.84} = 2.56$$

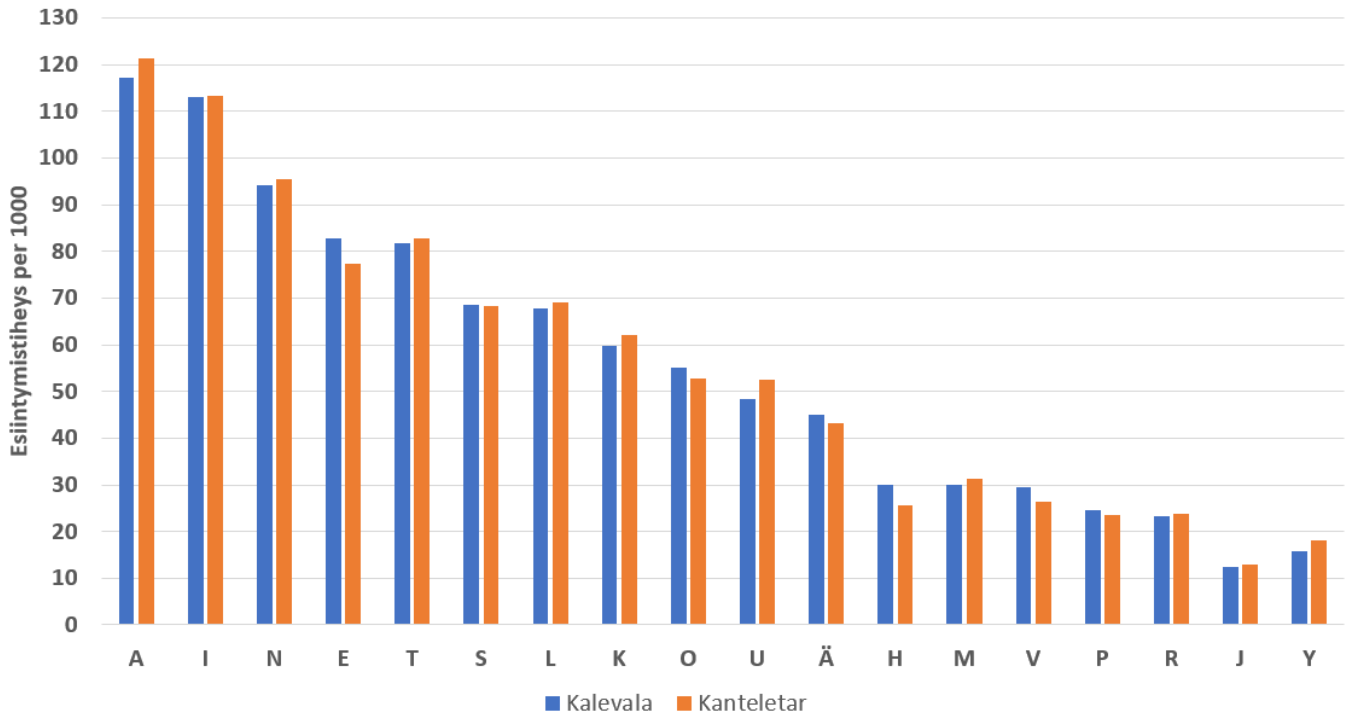
On helppo nähdä, että tämä lukema on sitä pienempi, mitä vähemmän toisiinsa verrattavat aineistot poikkeavat toisistaan.

Tehdään vastaava laskelma kaikille kirjoille ja saadaan taulukon 4 mukainen tulos, kalevalamitan mukaiseen järjestykseen asetettuna.

Listan kärjessä on Kanteletar, suorastaan ylivoimaisessa johdossa. Se ei ole millään tavoin hämmästyttävää, sillä onhan syyllinen sama henkilö, Paikkarin Elias. Lönnrotia voidaan syyttää sekä Kalevalan että Kantelettaren sisältämien runouksien keräämisestä. Kukaan ei häntä siihen käsenyt, mutta hän teki sen silti.

Kalevala ja Kanteletar ovat kirjainten esiintymistiheyksien suhteen hyvin yhteneväiset teokset. Se näkyy ehkä selvimmin kuvan 1 pylväskaaviosta. En muuten ole lukenut Kanteletartakaan. Vaikuttaa tämän

tutkimuksen perusteella huolestuttavan samanlaiselta teokselta kuin Kalevala. Ainoa ero on siinä, että Kanteletarta minulla ei (vielä) ole hyllyssäni.



Kuva 1. Kalevalan ja Kantelettaren kirjainten esiintymistiheydet. Viimeisenä olevat, Y-kirjaimella merkityt pylvää edustavat harvinaisimpien kirjainten Y, Ö, G, D, B, C, F, Q, W, X, Z ja Å koostetta.

Jos ei nyt Kanteletarkaan houkuttele, niin voisi harkita seuraavaksi Eino Leinon Helkavirsiä tai Anni Swanin Satuja lapsille luettavaksi? Täytyy ehkä kokeilla, vaikkakaan ei intoa puhkuen.

Suomenkielisessä aineistossa viimeiseksi asettuu teksti-tv:n sivusto. Ilmeinen selitys on siinä, että uutissivustolla varmasti esiintyy suhteellisen paljon vierasperäisiä paikan ja ihmisten nimiä ja näissä käytetään kirjaimia, joita Kalevalassa ei juurikaan nähdä. Mielenkiintoinen yksittäinen havainto on se, että YLE:ssä on ilmeisesti töissä paljon keski- tai peräpohjalaisia toimittajia. ”*Ei sitä hootakhan senthän joka paikhan laitetha*” eli H-kirjain todellakin esiintyy YLE-Teksti-TV:n tekstissä yllättävän harvoin verrattuna ainakin muihin tutkimuksessa olleisiin suomenkielisiin tekstilähteisiin.

Tutkimusaineistoon otin mukaan vertailun vuoksi kaksi englanninkielistä ja kaksi ruotsinkielistä kirjaa. Huomataan, että ne erottuvat selkeästi muusta aineistosta. Kirjainten jakauma on suomalaisessa tekstissä merkittävästi erilainen kuin englannin- tai ruotsinkielisessä tekstissä.

Suomen kieli on hyvin vokaalipitoinen kieli. Vokaalien ja konsonanttien keskinäinen suhde on aineistossa olleissa suomalaisissa teksteissä 48.0 % – 52.0 %, kun taas englanninkielisissä teksteissä se oli 40.3 % – 59.7 % ja ruotsinkielisissä teksteissä 36.4 % – 63.6 %.

Tämän pitkälti vokaali-konsonanttisuhteeseen perustuvan eron suomen ja ruotsin kielen välillä totesi Zacharias Topelius Maamme-kirjassaan (Topelius 1875) seuraavasti: ”*Ruotsin kieli ei ole niin pehmeätä, taipuisaa ja hienovivahteista kuin esim. suomen kieli, mutta se on miehuullista ja voimakasta, se kalskahtaa kuin teräs kalliioon.*”

Taulukko 4. Kirjojen ns. kalevalamitta

Kirja	Kalevalamitta $\chi^2$
Kanteletar	2.56
Helkavirsiä	16.11
Satuja lapsille luettavaksi	16.96
Seitsemän veljestä	21.72
Työmiehen vaimo	23.81
Punainen viiva	24.68
Vänrikki Stoolin tarinat	25.51
Putkinotko	25.72
Tulitikkuja lainaamassa	27.27
Juha	31.87
Mirdja	41.83
Teksti-tv 16.2.2022	41.85
Götiska rummen	1774.09
The Adventures of Sherlock Holmes	1827.12
Wuthering Heights	1875.78
Hemsöborna	1936.54

Niinpä niin. Onko sitten Kalevalan lukeminen muka miehuullista, vaikka se onkin kirjoitettu suomen kielellä? Onhan se. Muinaiset suomalaiset sankaritarustot ovat varmasti lukemisen arvoisia. Jotakin menneiden vuosituhansien takaisten esi-isien perinnöstä on talletettuna näihin kansalliseepoksemme säkeisiin. He suorastaan huutavat kumpujensa kätköistä minulle!

Onko Kalevala pehmeää, taipuisaa ja hienovivahteista? Johan nyt toki! Otan lakin päästäni ja jatkan kauan sitten pahasti kesken jääneen Kalevalan lukemista sivulta 203. *"Vaka vanha Väinämöinen, tietäjä iänikuinen, oli veistävä venoista, uutta purtta puhoava."*

Rakas lukijani. Kyllä minä siihen pystyn. Ihan varmasti. Usko pois.

Hyvää Kalevalanpäivää!

### Juhani Paananen

Matematiikan lehtori

Seinäjoen ammattikorkeakoulu

### Lähteet:

Aho, J. (1911): Juha, <https://www.gutenberg.org/ebooks/10863>

Brontë, E. (1847): Wuthering Heights, <https://www.gutenberg.org/ebooks/768>

Canth, M. (1885): Työmiehen vaimo, <https://www.gutenberg.org/ebooks/11296>

Conan Doyle, A. (1892): The Adventures of Sherlock Holmes, <https://www.gutenberg.org/ebooks/1661>

Kianto, I. (1909): Punainen viiva, <https://www.gutenberg.org/ebooks/54000>

Kivi, A. (1870): Seitsemän veljestä, <https://www.gutenberg.org/ebooks/11940>

Lassila, M. (1910): Tulitikkuja lainaamassa, <https://www.gutenberg.org/ebooks/10927>

Lehtonen, J. (1920): Putkinotko, <https://www.gutenberg.org/ebooks/39261>

Leino, E. (1916): Helkavirsiä, <https://www.gutenberg.org/ebooks/13129>

Lönnrot, E. (1835): Kalevala, <https://www.gutenberg.org/ebooks/7000>

Lönnrot, E. (1840): Kanteletar, <https://www.gutenberg.org/ebooks/7078>

Onerva, L. (1908): Mirdja, <https://www.gutenberg.org/ebooks/54306>

Pääkkönen, M. (1991): A:sta Ö:hön, suomen yleiskielen kirjaintilastoja, Kielikello 1/1991, s. 3, <https://jkorpela.fi/kielikello/kirjtil.html>

Runeberg, J.L. (1860): Vänrikki Stoolin tarinat, <https://www.gutenberg.org/ebooks/12688>

Strindberg, A. (1904): Götiska rummen: Släktöden från sekelslutet, <https://www.gutenberg.org/ebooks/48060>

Strindberg, A. (1887): Hemsöborna, <https://www.gutenberg.org/ebooks/30078>

Swan, A. (1901): Satuja lapsille luettavaksi, <https://www.gutenberg.org/ebooks/62936>

Teksti-TV 16.02.2022, <https://yle.fi/aihe/tekstiv>

Topelius, Z. (1875): Maamme-kirja, <https://www.gutenberg.org/ebooks/49600>