

KARL PAJUSALU JA URALILAISEN TYPOLOGISEN AINEISTON SYNTY

Outi Vesakoski

Tapasin Karl Pajusalun ensimmäistä kertaa osallistuessani viron kielen alkeiskurssille heti ensimmäisenä opiskeluvuoteni. Eipä aikaakaan, kun huomasin olevani järjestämässä Pajusalun vetämää opintoretkeä Saarenmaalle. Lehtorin innostus viron kieleen oli tarttuvaa, ja parin vuoden kuluttua olin Tartossa maantieteen kursseilla asuen Pajusaluilla yläkerran talonmiehenä. 20 vuotta myöhemmin palasimme yhteisille retkille, ja aloitimme uralilaisten kielten typologisen aineiston keruuhankkeen, jonka tulos – UraTyp eli uralilaisten kielten typologinen aineisto¹ – julkaistiin vuosi sitten. Nykyinen BEDLAN²-työryhmän yhteistyö Tarton yliopiston kanssa niin kielen, genetiikan kuin arkeologiankin saralla rakentuu sille perustalle, jonka Pajusalu välittömyydellään ja innostuneisuudellaan rakensi melkein 30 vuotta sitten.

UraTyp kuvaa 360 kielitypologisen piirteen esiintymistä uralilaisissa kielissä. Aineisto koostuu sekä ns. Grambank³-piirteistä (GB, N=195) että uralilaisille kielille tärkeistä piirteistä (UT, N=165). Käytännössä aineisto on 360 kysymystä siitä, että esiintyykö jokin piirre kielessä (kuva 1).

ID	Feature	Surgut Khanty	Hungarian	Komi- Zyrian	Meadow Mari	Pite Saami	Estonian
UT023	Can tense be expressed overtly on the negative marker?	0	0	1	1	1	0
UT078	Are there case categories only found in pronouns?	1	0	0	0	0	0
UT079	Does the numeral modification of a noun have an effect on its case selection?	0	0	0	0	0	1
UT080	Is the singular form of the noun also used with numerals bigger than 1?	1	1	1	1	0	1
UT105	Can 3Sg possessive suffix be used to mark definiteness?	0	0	1	1	0	0

Kuva 1. Esimerkkejä UT-piirteistä ja niiden esiintyminen muutamissa uralilaisissa kielissä. 1 = funktio on kielessä, 0 = sitä ei ole. Kuvan tekijä: Miina Norvik

¹ Uralic Areal Typology Online (uralic.clld.org).

² Biological Evolution and Diversification of LANguages (www.bedlan.net).

³ Grambank – a database of structural (typological) features of language (grambank.clld.org).

Tärkeässä roolissa yhteistyössä Turun ja Tarton välillä ovat olleet Karl Pajusalun järjestämät vuosittaiset seminaarit ja työpajat. Vuonna 2017 esittelimme ”Databases of the Uralic languages” – työpajassa BEDLAN-hankkeessa kerätyn uralilaisen typologisen aineiston sen hetkistä tilaa. Aineisto koostui 65 binäärisestä kielipiirteestä. Terhi Honkola oli analysoinut siihen saakka keräämämme aineiston populaatiogenetiikasta lainatulla klusterointimenetelmällä (esim. Syrjänen et al. 2016; Honkola et al. 2019) mutta tulos ei ollut tyydyttävä. Työ oli ollut hankalaa, ja koossa oli vain 16 kieltä, nekin heikosti katetusti. Piirteiden määrä ei todellakaan ollut riittävä kertomaan yhtään mistään. Menimme Tarton kokoukseen karvalakki kädessä anelemaan apua.

Miten tähän tilanteeseen oli päädytty? Lähtökohtana typologisen aineiston keruulle oli nuoren BEDLAN-työryhmän kokous Saaristomeren tutkimuslaitoksella Nauvossa vuonna 2010. Tällöin Michael Riebler ehdotti uralilaisten kielten typologisen aineiston keruuta uralilaisen perussanaston kognaatitietojen (UraLex, Syrjänen et al. 2021, de Heer et al., in print) ja Suomen murreatlaksen oheen. Ajattelin, että ei kyllä jaksaisi, eikä kyllä kykenekään, mutta kohtapa aihe oli Unni-Päivä Leinon Koneen Säätiön hankkeessa ja BEDLAN:iin palkattiin Jyri Lehtinen suunnittelemaan piirrelistaa. Lehtinen lähti hakemaan oppia Nijmegenistä, psykolingvistiikan Max Planck Instituutista (MPI) Michael Dunnilta. Tämä oli työstänyt Stephen Levinsonin masinoimassa koalitiossa kielen rakennepiirteistä listoja, jotka olisivat sopivia kielihistorian tutkimiseen. Ns. Nijmegenin listasta valitsimme 65 kysymystä, jotka kysyivät, oliko kielessä jokin tietty piirre. Vastaukset kysymyksiin etsittiin kieliopista ja kielten kuvauksista. Piirteet kattavaa kirjallisuutta ei kuitenkaan ollut tarpeeksi tarjolla.

Samaan aikaan toisaalla Max Planckin uudessa instituutissa Sciences for Human Historyssä (MPI-SHH), Russell Gray halusi alkaa uudistaa WALSiä. Pohjaksi otettiin Nijmegenissä tehty piirrelistaus, jota alettiin puida isossa työryhmässä vuodesta 2014. Minä olin mukana uralilaisen yritelmämme takia. Tarkoitus oli, että uralilaiset kielet kerättäisiin samaan maailman kielten aineistoon. MPI-SHH rakensi Grambank-listan (GB-piirteet), joka sisälsi 195 kysymystä kielen rakenteesta, sekä kehitti systemaattisen järjestelmän, jolla kieliopista ja kielikuvauksista löydettäisiin binäärinen on/ei -vastaus kysymyksiin. Kerroin tästä Tarton kokouksessa samalla kun raportoin omista tuloksistamme ja kokemuksistamme. Ehkä GB-piirrekysymyksiin vastaaminen uralilaisten kielten osalta olisi ratkaisu.

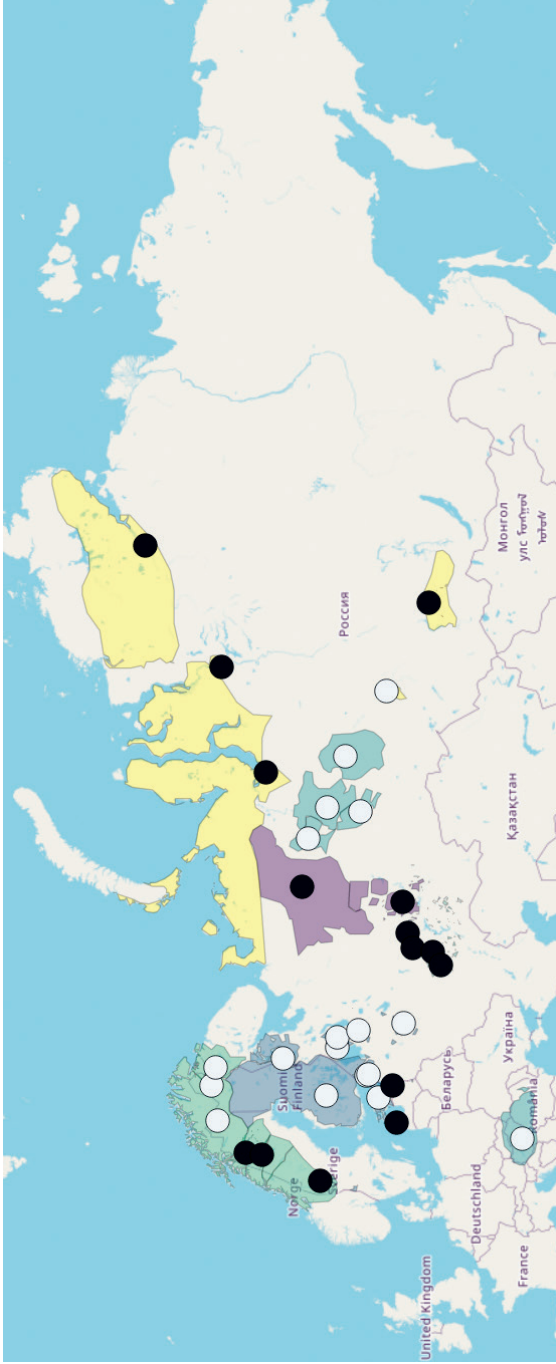
Nyt tähdet alkoivat asettua oikeisiin aseisiin. Karl Pajusalu tutustutti Miina Norvikin BEDLAN-työryhmään, ja Turun yliopisto avasi sopivasti haun ”strategisten tutkimusaineistojen keräämiseksi”. Haimme rahoitusta hankkeelle ”Kipot ja kielet”. Päähakijana oli prof. Päivi Onkamo, ja mukana olivat prof. Sirkka Saarinen Turun yliopistosta ja Karl Pajusalu Tartosta ja koordinaattoriksi ryhtyvä Miina Norvik (hankkeen kokonaiskuvaus: Moilanen et al. 2021).

Hanke alkoi sillä, että Miina Norvik, Karl Pajusalu, Helle Metslang, Gerson Klumpp ja Eva Saar istuivat alas kerran jos toisenkin ja koostivat GB-piirteiden ohien uralilaisille kielille relevanttien UT-piirteiden listan. Tarkoitus oli saada aineistoon mukaan sellaisia typologisia piirteitä, joiden avulla uralilaisia kieliä voisi verrata toisiinsa – GB-piirteet oli valittu kielikuntien väliseen vertailuun. GB-aineiston keräämiseen palkattiin Richard Kowalik Tukholman yliopistosta ja UT-aineiston keräämiseen Minerva Piha Turun yliopistosta sekä Eva Saar Tarton yliopistosta. Työ oli mahdollista nyt, kun pystyimme käyttämään Grambank-työryhmän rakentamaa systemaattista aineistonkeräysmenetelmää. Toisaalta uralilaisia kieliä varten meidän piti sopia MPI:n kanssa, että voimme käyttää kielentutkijoita informantteina niissä tapauksissa, kun kielioppeja ei ole tarjolla.

Grambank-työryhmässä päätettiin, että GB-aineistoon kerätään vain tieto siitä, esiintyykö piirre kielessä vai ei. Esimerkkejä ei kerätty, sillä ensimmäisessä vaiheessa tarkoitus oli saada kerättyä mahdollisimman monen kielen binäärinen tieto. Tarton työryhmä kuitenkin halusi esimerkit mukaan UT-piirteisiin. Tämä tuotti lisää työtä, mutta mahdollisti aineiston syvemmän käytön.

Uralilainen iisakinkirkko valmistui 2022, jolloin saimme julkaistua UraTyp-aineiston (Norvik et al. 2021) ja siihen liittyvän artikkelin (Norvik et al. 2022). Koko Grambank-aineisto julkaistaan samoihin aikoihin, kun tämä kirja menee painoon (Skirgård et al. 2023), eli tätä lukiessasi käytöissäsi on 2500 kielen typologinen aineisto, josta löytyy myös 30 uralilaisista kieltä. UraTyp-aineisto löytyy jäädytettynä versiona Zenodo-repositorystä, sekä helposti lähestyttävästi Uralic Areal Typology Online käyttöliittymästä (kuva 2). Käyttöliittymän oli tehnyt valmiiksi Robert Forkel (Max Planck Institute for Evolutionary Anthropology) jo osana Grambank-aineistojen esilletuontia.

Aineiston keruun edetessä työryhmälle kävi selväksi, että aineisto ei koskaan tule olemaan valmis. Sitä voi laajentaa kattamaan enemmän kieliä



Kuva 2. Näkymä UraTypen käyttöliittymästä Uralic Areal Typology Online (uralic.clld.org): UT023 piirteen esiintyminen. Musta pallo = funktio on kielessä, valkea = sitä ei ole. Tausstakartat Rantanen et al. 2022. Kuvakaappaus 16.3.2022

tai kielten sisäistä variaatiota, piirteitä voi lisätä, binäärisen funktio-tiedon lisäksi voidaan esimerkiksi kerätä esimerkkeihin perustuen kognaatitieto joidenkin piirteiden muodoista. UraTyp-aineiston ensimmäisen version julkaisun jälkeen Tarton työryhmä on Rõukin perheen perintörahaston avulla jatkanutkin aineiston kehittämistä, ja seuraavaan UraTypen versioon tulee mukaan esimerkit myös GB-piirteistä. Julkaisun jälkeen Sven-Erik Soosaar, Miina Norvik ja John Huisman ovat lisäksi laajentaneet UT-aineiston keräämistä myös naapurikielistä. Aineistoa ylläpidetään ja uudistetaan GitHub-repositoryssä, jossa sitä voi muokata periaatteessa helposti, ja eri versioista säilyy tieto. Uusia versioita tullaan julkaisemaan mm. käyttöliittymässä.

UraTyp-artikkeli oli ensimmäinen yhteinen käsikirjoitukseni Karl Pajusalun kanssa, ja sitä tehdessämme tiesin mitä teemme ja mihin matkaamme. Toisin oli yhteistyömme alkumetreillä noin vuonna 1993, kun lehtori Pajusalu rekrytoi minut auttamaan Saarenmaan matkan järjestämisessä ja tekemään kilpailutusta bussifirmoille. Olin pihalla kuin jänis kaikesta käytännön säädöstä – mutta pääsimme perille ja matkan jälkeen suunnittelimme jo seuraavaa. Niin on nytkin – UraTyp on valmis, mutta seuraavaa versiota tehdään jo. Professori Pajusalun kyydissä on hyvä matkustaa, sillä kuskilla on visio, muutosvoimaa ja kyky saada ryhmä liikkumaan yhteen suuntaan.

Lähteet

De Heer, Mervi, Rogier Blokland, Michael Dunn, Outi Vesakoski [in print].

Loanwords in basic vocabulary as an indicator of borrowing profiles. – *Journal of Language Contact*.

Honkola, Terhi, Jenni Santaharju, Kaj Syrjänen, Karl Pajusalu 2019.

Clustering lexical variation of Finnic languages based on *Atlas Linguarum Fennicarum*. – *Linguistica Uralica* 55 (3), 161–184. <https://dx.doi.org/10.3176/lu.2019.3.01>.

Moilanen, Ulla, Petro Pesonen, Miina Norvik, Jarkko Saipio, Outi Vesakoski,

Vesa Immonen, Päivi Onkamo 2021. New Tools for studying Finnish archaeology and Uralic Languages. – *Antiquity* 95 (383), 1–6. <https://doi.org/10.15184/aqy.2021.113>.

Norvik, Miina, Yingqi Jing, Michael Dunn, Robert Forkel, Terhi Honkola,

Gerson Klumpp, Richard Kowalik, Helle Metslang, Karl Pajusalu, Minerva Piha, Eva Saar, Sirkka Saarinen, Outi Vesakoski 2021. Uralic Typological database – UraTyp [Data set]. 10.5281/zenodo.6392555.

- Norvik, Miina, Yingqi Jing, Michael Dunn, Robert Forkel, Terhi Honkola, Gerson Klumpp, Richard Kowalik, Helle Metslang, Karl Pajusalu, Minerva Piha, Eva Saar, Sirkka Saarinen, Outi Vesakoski 2022.** Uralic typology in the light of a new comprehensive data set. – *Journal of Uralic Linguistics* 1 (1), 4–42. <https://doi.org/10.1075/jul.00002.nor>.
- Rantanen, Timo, Harri Tolvanen, Meeli Roose, Jussi Ylikoski, Outi Vesakoski 2022.** Best practices for spatial language data harmonization, sharing and map creation – A case study of Uralic. – *PLoS ONE* 17 (6), 1–19. <https://doi.org/10.1371/journal.pone.0269648>.
- Skirgård, Hedvig, Hannah J. Haynie, Damián E. Blasi et al. 2023.** Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of languages loss. – *Science Advances* 9 (16), 1–15. <https://doi.org/10.1126/sciadv.adg6175>.
- Syrjänen, Kaj, Terhi Honkola, Jyri Lehtinen, Antti Leino, Outi Vesakoski 2016.** Applying population genetic approaches within languages: Finnish dialects as linguistic populations. – *Language Dynamics and Change* 6, 235–283.
- Syrjänen, Kaj, Luke Maurits, Unni Leino, Terhi Honkola, Jadranka Rota, Outi Vesakoski 2021.** Crouching TIGER, hidden structure: Exploring the nature of linguistic data using TIGER values. – *Journal of Language Evolution* 6 (2), 99–118. <https://doi.org/10.1093/jole/lzab004>.

Outi Vesakoski: Karl Pajusalu and the creation of the Uralic typological database

In the year 2022 we published the Uralic Typological Dataset (UraTyp, <https://uralic.clld.org>). The BEDLAN working group had started working on this initiative in Finland many years earlier, but without success. Things only got serious after Finnish *sisu* was combined with the talents of prof Karl Pajusalu and colleagues, and with the Max Planck Institute’s methodological developments. This is a “behind the scenes” story on Finnish-Estonian co-operation in tabulating Uralic typology to build a comparative database for Uralic and global studies – and how this all is related to tendering bus companies to get to Saaremaa.

Outi Vesakoski: Karl Pajusalu ja uurali kiili tüpoloogilidsi andmidõ sünd

Aastagal 2022 mi ilmudi Uurali kiili tüpoloogilidsõ andmõvõrgo (UraTyp, <https://uralic.clld.org>). BEDLANi tüürühm oll' naanuq tuu tüüga pääle jo pall'o aastit inne tuud, a ilma suurõmbidõ tulõmiildaq. Aśaq naksiq tõsit-sõmpa liikma päält tuud, ku Soomõ *sisu* panti kokko prohvosri Karl Pajusalu ja tä tüüseldsiliisi anniga ni Max Plancki Instituudi metodoloogilidsi edenemiisiga. Seo om „kaardinidõtagonõ“ lugu Soomõ-Eesti kuuntüüst uurali kiili tüpologiatabõlidõ kokkosäädmise man, et luvvaq võrdõlõjat teedüskoko uurali ja üleilmaliidsi uurmiisi jaos – ja tuust, kuis tuu kõik om kõüdet bussifirmadõ pakmiisiga Saarõmaalõ saamisõs.