

# Kielten ja murteiden evoluutio ja laskennalliset tekniikat

KAJ SYRJÄNEN

## Väitöksenalkajaisitelmä Tampereen yliopistossa 6. elokuuta 2021

Ajatus siitä, että kieliä ja biologisia lajeja voisi samaistaa toisiinsa, kuulostaa kenties hieman absurdilta. Tämä idea on kuitenkin esitetty jo yli sata vuotta sitten, ja senkin jälkeen useaan otteeseen eri yhteyksissä. Sitä ovat nostaneet esille muun muassa evoluutioteorian isä Charles Darwin ja kielten sukupuuteorian kehittänyt August Schleicher. Yhteistä näillä kahdella oli muun muassa se, että molemmat käyttivät puukuvaajia historian esittämiseen (Atkinson & Gray 2005). Heidän teksteissään myös ajoittain nostettiin esille kielihistorian ja biologisen evoluution samankaltaisuutta.

Millaista samankaltaisuutta näiden välillä sitten on? Ehdotettuja samankaltaisuuksia löytyy useampia (ks. esim. Pagel 2009), joskin kolme on mielestäni muita keskeisempiä. Ensinnäkin sekä kieliin että biologisiin lajeihin liittyy tiedon välittymisen ajanjaksosta toiseen yksilöiden kautta. Kielten tapauksessa kieliaines, kuten esimerkiksi sanasto, äänteet ja kieliopilliset rakenteet, välittyvät kielenkäyttäjältä toiselle kielenoppimisprosessin ja ylipäätään viestinnän kautta. Biologiassa puolestaan yksilön geneettinen perimä koostuu yhdistelmästä tämän vanhempien perimää; jokainen jälkeläinen säilyttää osan vanhempiensa geneettisestä informaatiosta. Tiedonvälitysprosessi on funktionaalisesti hyvinkin erilaista kielillä ja biologisilla lajeilla, mutta molemmille yhteistä on informaation välittyminen ja säilyminen ajanjaksosta toiseen yksilöiden kautta. Yleisessä evoluutioteoriassa tätä kieliä ja biologisia lajeja yhdistävää ominaisuutta kutsutaan *periytymisen periaatteeksi*.

Voidaan ajatella, että kielellinen tietämyksemme on eräällä tavalla kokemustemme summa – siinä yhdistyvät eri lähteistä oppimamme ja omaksumamme asiat. Tämä kielellinen tietämys on kokonaisvaltaisesti katsottuna uniikki jokaiselle puhujalle. Vaikka puhuisimmekin keskenämme samaa kieltä, jokainen meistä puhuu sitä hieman omalla tavallaan ja suosii esimerkiksi tietyille asioille tiettyjä nimityksiä. Jokainen periaatteessa samaa kieltä puhuva henkilö ei välttämättä tunnista kaikkia tietyn yksilön käyttämistä sanoista, ainakaan suoralta kädeltä.

Tämänkin ominaisuuden osalta kielitietämys muistuttaa jossain määrin biologista perimää, joka on uniikki jokaiselle yksilölle, vaikka nämä kuuluisivat samaan lajiin. Pääasiassa kaikki yksilöt omaavat funktionaalisesti samaa asiaa ajavat osat, kuten biologisen perimän tapauksessa geenit tai kielten osalta kuhunkin merkitykseen assosioituvat sanat. Näissä osissa kuitenkin on aina yksilötason eroja, kuten vaikkapa saman geenin eri variantit eli alleelit tai eri murretaustaisten puhujien vaihtelevat ilmaukset kullekin merkitykselle. Yleisessä evoluutioteoriassa tätä ominaisuutta kutsutaan *muuntelun periaatteeksi*.

Kolmanneksi sekä kielissä että biologisissa lajeissa kielellinen tieto ei välity täysin muuttumattomana ajanjaksosta toiseen. Osa sanoista tai kielirakenteista jää omaksuttamatta, tai ne omaksutaan harvemmin tai erilaisissa muodoissa kuin missä ne lähtökohtaisesti olivat. Samalla kieli ei pysy yhtenäisenä aikakaudesta toiseen, vaan se mukautuu jatkuvasti vastaamaan niitä kommunikatiivisia tarpeita, joita sen puhujilla kulloinkin on. Nämä sisäiset ja ulkoiset muospaineet, jotka aiheuttavat niin suunnattua kuin sattumanvaraista muutosta, voi samaistaa ilmiöön, jota yleisen evoluutioteorian terminologiassa kutsutaan *valinnan periaatteeksi*.

Edellä mainitut kolme melko yleistason käsitettä – *periytyvyys, muuntelu ja valinta* – ovat evoluution kolme pääperiaatetta. Hyvin yleisellä tasolla kulttuuri-ilmiöiden kehitystä, kuten vaikkapa kielenmuutosta, voidaan myös käsitteellistää näiden periaatteiden kautta, kuten edellä esitettiin. Sen myötä voidaan jossain määrin ajatella, että kielikin on evolutiivinen ilmiö.

Olettaen, että hyväksymme ajatuksen kielestä evolutiivisena ilmiönä, herää kysymys, mitä hyötyä tästä on. Yksi ajatuksen tarjoamista hyödyistä on se, että voimme evoluutioteorian kautta pyrkiä rakentamaan kattavampaa kuvaa kielenmuutosprosessista katsomalla, mitkä osat kielenmuutoksesta samaistuvat minkäkinlaisiin evoluution käsitteisiin tai mekanismeihin. Tätä voidaan tehdä kokonaisvaltaisen, evoluution päälle rakennetun kielenmuutosteorian avulla, kuten esimerkiksi William Croft (2000, 2008) on pyrkinyt tekemään, tai suppeammin, yksittäisten käsitteiden samankaltaisuuksiin keskittyen, joita on esitellyt muun muassa Pagel (2009). Toisenlaisen tulokulman tarjoavat evoluutiobiologiassa kehitetyt laskennalliset tekniikat, joiden soveltuvuutta kielitieteellisen ja muunlaisen kulttuuriaineiston analysoimiseen on alettu tutkia viime vuosikymmeninä. Tähän aihepiiriin myös oma väitöskirjatutkimukseni pääosin keskittyy.

Väitöstutkimukseni pääpainopiste on evoluutiobiologiasta omaksutun metodologian testaamisessa kieliaineistolla. Kokonaisuus koostuu neljästä tutkimusartikkelista, jotka käsittelevät kahta isompaa aihepiiriä: sukukielten välisiä yhteyksiä, joita omassa työssäni edustivat suomelle sukua olevat uralilaiset kielet (Syrjänen, Honkola, Korhonen, Lehtinen, Vesakoski & Wahlberg 2013; Lehtinen, Honkola, Korhonen, Syrjänen, Wahlberg, & Vesakoski 2014), ja kielensisäistä vaihtelua, jota omassa tutkimuksessani edustivat suomen murteet (Syrjänen, Honkola, Lehtinen, Leino & Vesakoski 2016). Osassa tutkimusta käytettiin myös keinotekoisesti simuloitua kieliaineistoa metodologian koestamisessa (Syrjänen, Maurits, Leino, Honkola, Rota & Vesakoski 2021).

Lajien välisiä suhteita tutkitaan evoluutiobiologiassa nykyisin *laskennallisen fylogeneetiikan* avulla. Fylogeneettisten tekniikoiden avulla esimerkiksi DNA-aineistosta saadaan laskennallisesti rakennettua historiasta tai samankaltaisuudesta kertovia

luokitteluja, kuten vaikkapa fylogeneettisia puita tai verkostoja. Tekniikoita alettiin soveltaa ja adoptoida kielitieteeseen parikymmentä vuotta sitten muun muassa indoeurooppalaisten ja austronesialaisten kielten tutkimuksessa, ja niitä on sittemmin sovellettu moniin muihinkin kielikuntiin (Greenhill, Heggarty & Gray 2020).

Omassa tutkimuksessani lähestyin uralilaisia kieliä fylogeneettisten tekniikoiden kautta. Tavoitteena oli sekä valaista näiden kielten välisiä suhteita että perehtyä fylogeneettisiin tekniikoihin ja niiden toimivuuteen kieliaineiston kanssa. Käytännössä tutkimuksessa DNA-aineiston sijasta fylogeneettisille ohjelmistoille syötettiin kieliaineistoa, joka kuvasti uralilaisten kielten välisiä historiallisia suhteita. Tässä käytetty kieliaineisto, UraLex, on kerätty osana tutkimustyötä ja julkaistu *Creative Commons*-lisenssin alla. Aineiston avulla oli mahdollista rakentaa fylogeneettisiä puita ja verkostoja näistä kielistä ja alkaa rakentamaan laskennallista kuvaa tästä kielikunnasta. Tähän aiheeseen keskittyivät erityisesti väitöskirjan kaksi ensimmäistä osatyötä (Syrjänen ym. 2013; Lehtinen ym. 2014).

Lajien sisäisiä suhteita tutkitaan puolestaan *populaatiogenetiikan* menetelmillä. Nämä laskennalliset menetelmät auttavat hahmottamaan samankaltaisten yksilöiden joukkoja eli populaatioita isosta joukosta yksilöitä ja siten rakentamaan kuvaa lajin sisäisestä evoluutiosta. Tekniikoita sovellettiin väitöskirjani kolmannessa osatyössä (Syrjänen ym. 2016), joka keskittyi laskennalliseen murrettutkimukseen. Aineistona toimi digitoitu versio Lauri Kettusen keräämästä murrekartastosta (Kettunen 1940), joka kokonaisuudessaan koostuu 213 karttalahdestä, jotka taltioivat murrepiirteitä Suomessa 1900-luvun alusta. Murrekartaston keruupisteitä, käytännössä siis Suomen eri pitäjiä, joista Lauri Kettunen aineistoa keräsi, samaistettiin populaatiogenetiikan yksilöihin, jotka puolestaan laskennallisesti luokiteltiin populaatioiksi eli tässä tapauksessa murrealueiksi. Väitöskirjani kolmannessa osatyössä näitä menetelmiä sovellettiin ensimmäistä kertaa murreaineistoon. Tätä samaa tekniikkaa on näiden tutkimusten jälkeen sovellettu myös itämerensuomalaisten kielten variaatiota sisältävään *Atlas Linguarum Fennicarum* -aineistoon (Honkola, Santaharju, Syrjänen & Pajusalu 2019), josta tämä tutkimus ei ole osa väitöskirjaani.

Tutkimuksen viimeisessä osatyössä (Syrjänen ym. 2021) palasin takaisin laskennallisen kielten luokittelun ongelmiin. Tällä kertaa keskiössä tosin ei ollut uralilaisen kielikunnan rakenne, vaan kielikuntien rakenne yleisemmin ja eräs siihen liittyvä haaste. Vaikka voimmekin rakentaa sukupuun periaatteessa mistä tahansa joukosta sukukieliä, tämä ei välttämättä ole luontevin tapa esittää kielten historiallisia suhteita toisiinsa, sillä kielipiirteet eivät ainoastaan periydy vaan myös lainautuvat eri kielten välillä. Tätä varten onkin hyödyllistä olla tapoja kysyä aineistolta laskennallisesti, kuinka puumainen se on, ennen kuin siitä rakentaa puun. Tähän kysymykseen tutkimukseni neljäs osatyö keskittyi. Työssä tutkin fylogeneettista TIGER-algoritmia ja sen kykyä mitata kieliaineiston puumaisuutta. Tässä työssä uralilaiset kielet ovat sivuroolissa, sillä tutkimus oli pääasiassa metodologinen kontribuutio.

Evoluutiobiologian tekniikat oikein sovellettuna tarjoavat tehokkaita työkaluja mallintaa myös kielivariaation ja kielihistorian ilmiöitä. Tekniikat ovat jatkuvasti kehittyviä ja niiden soveltuvuus ja ymmärryksemme niistä tulee lisääntymään myös tulevana vuosina.

## Lähteet

- ATKINSON, QUENTIN – GRAY, RUSSELL D. 2005: Curious parallels and curious connections. Phylogenetic thinking in biology and historical linguistics. – *Systematic Biology* 54 (4) s. 513–526. <https://doi.org/10.1080/10635150590950317>.
- CROFT, WILLIAM 2000: *Explaining language change. An evolutionary approach*. Essex: Pearson Education Limited.
- 2008: Evolutionary linguistics. – *Annual Review of Anthropology* 37 s. 219–234. <https://doi.org/10.1146/annurev.anthro.37.081407.085156>.
- GREENHILL, SIMON J. – HEGGARTY, PAUL – GRAY, RUSSELL D. 2020: Bayesian phylo-linguistics. – Richard D. Janda, Brian D. Joseph & Barbara S. Vance (toim.), *The Handbook of Historical Linguistics. Volume 2* s. 226–253. Hoboken: Wiley. <https://doi.org/10.1002/9781118732168.ch11>.
- HONKOLA, TERHI – SANTAHARJU, JENNI – SYRJÄNEN, KAJ – PAJUSALU, KARL 2019: Clustering lexical variation of Finnic languages based on Atlas Linguarum Fennicarum. – *Linguistica Uralica* 55 (3) s. 161–184. <https://doi.org/10.3176/lu.2019.3.01>.
- KETTUNEN, LAURI 1940: *Suomen murteet. III A. Murrekartasto*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- LEHTINEN, JYRI – HONKOLA, TERHI – KORHONEN, KALLE – SYRJÄNEN, KAJ – WAHLBERG, NIKLAS – VESAKOSKI, OUTI 2014: Behind family trees. Secondary connections in Uralic language networks. – *Language Dynamics and Change* 4 (2) s. 189–221. <https://doi.org/10.1163/22105832-00402007>.
- PAGEL, MARK 2009: Human language as a culturally transmitted replicator. – *Nature Reviews Genetics* 10 (6) s. 405–415. <https://doi.org/10.1038/nrg2560>.
- SYRJÄNEN, KAJ – HONKOLA, TERHI – LEHTINEN, JYRI – LEINO, ANTTI – VESAKOSKI, OUTI 2016: Applying population genetic approaches within languages. Finnish dialects as linguistic populations. – *Language Dynamics and Change* 6 (2) s. 235–283. <https://doi.org/10.1163/22105832-00602002>.
- SYRJÄNEN, KAJ – HONKOLA, TERHI – KORHONEN, KALLE – LEHTINEN, JYRI – VESAKOSKI, OUTI – WAHLBERG, NIKLAS 2013: Shedding more light on language classification using basic vocabularies and phylogenetic methods: A case study of Uralic. – *Diachronica* 30 (3) s. 323–352. <https://doi.org/10.1075/dia.30.3.02syj>.
- SYRJÄNEN, KAJ – MAURITS, LUKE – LEINO, UNNI – HONKOLA, TERHI – ROTA, JADRANKA – VESAKOSKI, OUTI 2021: Crouching TIGER, hidden structure. Exploring the nature of linguistic data using TIGER values. – *Journal of Language Evolution* 6 (2) s. 99–118. <https://doi.org/10.1093/jole/lzab004>.

**Kaj Syrjänen: Quantitative Language Evolution. Case studies in Finnish dialects and Uralic languages.** Tampere University Dissertations 433. Tampere: Tampereen yliopisto 2021. Kirja on luettavissa osoitteessa: <http://urn.fi/URN:ISBN:978-952-03-2004-1>.

Kirjoittajan yhteystiedot:  
etunimi.sukunimi@tuni.fi