

# Karjalan kielen koodinvuorovaikutus

SUSANNA TAVI

## Väitöksenalkajaisesityelmä Itä-Suomen yliopistossa 17. helmikuuta 2023

Sanasto on yksi kielen tutkituimpia osa-alueita. Se johtunee siitä, että sanastoa voidaan pitää kommunikaation ytimenä. Vähällä määrällä sanoja voikin päästä pitkälle, mitä havainnollistaa esimerkiksi puhumaan opettelevien lasten ja uutta kieltä opettelevien aikuisten kielenkäyttö. Erityisesti sanastoa on tutkittu sanojen etymologian eli alkuperän näkökulmasta: Mistä ja miten sanat kieleen tulevat? Ovatko ne kenties omaperäistä kehitystä vai lainoja toisista kielistä? Aihetta on tutkittu niin runsaasti, että viime aikoina suosiossa ovatkin olleet aivan toisenlaiset tutkimuskysymykset.

Toisenlaiset tutkimuskysymykset ovat olleet suosiossa myös kontaktilingvistiikassa eli kielikontaktien ja kielten välisen vaikutuksen tutkimuksessa. Kontaktilingvistiikan suosion lähdettyä nousuun 1950- ja 1960-luvulla tutkimuskysymykset monikielisydestä ovat yleistyneet (mm. Haugen 1950 ja Weinreich 1968 [1953]). Koodinvaihto eli karkeasti kuvattuna kahden kielen aineisten vaihtelu yhden lausuman sisällä sekä erilaisten rakenteellisten piirteiden lainautuminen sanaston sijaan ovat olleet kontaktilingvistisen tutkimuksen ytimessä. Sanaston tarkastelu on jäänyt kontaktilingvistiikassa yllättävän vähäiselle huomiolle, vaikka useiden kielten sanastoissa esiintyykin kontaktilähtöistä variaatiota eli vaihtelua. Näin on myös suomen lähimmässä sukukielessä karjalassa. Karjala on uhanalainen vähemmistökieli, jota puhutaan Suomessa ja Venäjällä. Karjalassa kontaktilähtöinen vaihtelu ilmenee esimerkiksi synonyymeinä, joista Natalia Giloeva on maininnut blogikirjoituksessaan (Giloeva 2015) merkitysvastineet *rikas* ~ *bohattu* 'rikas', *kirju* ~ *kniigu* 'kirja' ja *järjestiä* ~ *organizuija* 'järjestää'.

Sanojen kontaktiperäistä synonyymista käyttöä ei ole tarkasteltu kokonaisvaltaisesta näkökulmasta. Ei siis ole tarkasteltu, kumpi näistä synonyymeistä valitaan kulloinkin ja millä perusteella. Halu tietää vastaus näihin kysymyksiin onkin ajanut minut tekemään väitöstutkimukseni. Väitöstutkimukseni on artikkelimuotoinen, ja sen neljässä artikkelissa olen käsitellyt eri näkökulmista karjalan kielen sanastoa kontaktissa suomen ja venäjän kieliin. Kahdessa ensimmäisessä artikkelissa (Tavi 2018 ja Tavi & Tavi 2019) tarkastelin Suomessa 1960- ja 1970-luvuilla puhuttuja rajakarjalaismurteita. Kol-

mannen artikkelin (Tavi & Tavi 2022) tutkimuskohteena olivat 2010-luvulla Venäjällä Karjalan Tasavallassa käytetyt vianan- ja livvinkarjalan kirjakielet ja neljännessä artikkelissa (Tavi 2022) aineisto koostui vuonna 2018 muistikävelymenetelmällä tallentamistani tverinkarjalaisten haastatteluista.

Tutkimukseni tarkoituksena on ollut yhdistää kriittisesti uhanalaisen karjalan kielen tutkimuksen hajanaisia näkökulmia siten, että sanaston kontakti-ilmiöt sijoittuvat jatkumolle aina Suomesta Sisä-Venäjälle Tveriin saakka. Aineisto koostuu sekä puhutusta että kirjoitetusta kielestä. Vaikka usein käsitystämme kirjoitetusta kielestä rajoitavatkin tiukat kirjakieltä määrittävät normit, molemmissa tehdään valintoja sanojen välillä. Se, mitä sanaa päädytään käyttämään tietoisesti tai tiedostamatta, vaikuttaa siihen, millaiseksi kieli kehittyy. Tällaisen huomion on tehnyt myös eräs tverinkarjalan puhuja, jota haastattelin vuonna 2018 Stuanun kylässä (ks. esimerkkiä 1).

- (1) No a pikkuzesta šuat ka mie muissan vielä oldih miän kyläššä ennen vie oldih, oli akkane. Miula pikkarane konza olin pieni šaneli šuarnasii. Vielä ei ma- ei hän ma- mahtannun hormaksi ni yhtä šanua.

Ni yhtä šanua?

Ni yhtä. Kaikin karielaksi pagiži vain čisto ka-. A šiidä šua kai jo vaihtuačči kaikki yhessä zivilizačii. Tuli- tuldihni šanat. Toizet. Kieleh näh. Zentäh vaihtuačči kieli. Äijäzelti vaihtuačči. A kun ennen oli so- čut karielane.

Puhuja mainitsee, että kieleen on tullut toiset sanat samalla, kun puhujat ovat alkaneet puhua karjalan lisäksi venäjää (*hormaksi*). Ketkä siis ovat niitä, jotka tuovat näitä toisia sanoja kieleen? Sanotaan, että monikielisillä yksilöillä on tärkeä rooli uusien innovaatioiden välittämisessä (mm. Kovács 2009: 44). Karjalan kielen tämänhetkinen tilanne on sellainen, että kaikki puhujat ovat kaksi- tai monikielisiä. He puhuvat asuinvaltiostaan riippuen karjalan lisäksi joko suomea tai venäjää. Monet etenkin kieliaktivisteina toimivat puhujat osaavat kaikkia näistä kielistä (Kuusi, Riionhiemi & Koehmainen 2022: 133–134, 146). Tällaisessa tilanteessa voidaan puhua kielellisistä resursseista eli varannoista: kaksikielisillä, tässä tapauksessa karjalaisilla, on mahdollisuus hyödyntää keskustelussa resursseja karjalan lisäksi vähintään yhdestä muustakin kielestä. Pitkään tällaiset resurssit on nähty negatiivisina ja omaa kieltä vahingoittavina elementteinä. Sittemmin on kuitenkin huomattu, että niiden käyttö vaatii erityistä taitoa ja ne ovat lisäresurssi vanhan omaperäisen tuhoamisen sijaan (Sarhimaa 1999: 197–198). Karjalaisen kielenkäytön luonteeseen kuuluvan kaksikielisyyden ymmärtäminen on noussut tärkeäksi näkökulmaksi myös karjalan kielen sanaston kontakti-ilmiöistä variaatiota tarkastelevassa väitöstutkimuksessani.

Kun aloin tutkia rajakarjalaismurteita pro gradu -työssäni, huomasin, että karjalasta on olemassa runsaasti arkistoitua aineistoa, mutta tutkimuksissa siitä hyödynnetään usein vain pieniä osia. Suuntasin ajatukseni kohti sellaisia menetelmiä, joiden avulla tutkimukseen voisi ottaa mukaan mahdollisimman paljon aineistoa. Sopivanlaisen viitekehysten luomisesta tulikin yksi väitöskirjani tavoitteista. Muita tavoitteita

ovat valitsemani teoreettisen viitekehyksen, koodinkopioimisen viitekehyksen, tarkastelu karjalaisessa kontekstissa sekä suomen ja venäjän kontaktivaikutuksen tarkastelu ja vertailu karjalan kielen sanastossa.

Moni aineisto on sellaisessa muodossa, että sitä ei ole voinut käsitellä helposti koineellisesti. Aineiston läpikäyminen on ollut käsityötä ja vaatinut siten lukemattomia tunteja tutkijoilta. Ratkaisuni aineiston käsittelyn helpottamiseksi kuulostaa ehkä jopa rujolta: laitoin aineiston korpusohjelmiin ikään kuin väkisin, ja katsoin, mitä sille tulisi tehdä, jotta ohjelma tunnistaa sanat. Seuraava lista havainnollistaa vaihteita, joita ehdotan annotoimattoman aineiston käsittelemiseksi.

1. Käsittele aineisto koneluettavaan muotoon: poista tekstistä muun muassa tärkeitä tekstin yksinkertaistamiseksi, käytä suomea apukielenä.
2. Käytä korpuslingvistisiä työkaluja: luo esimerkiksi sanalistoja ja käytä regular expressions -komentoja sekä skriptejä listojen läpikäymiseksi.
3. Järjestele ja kvantifioi aineisto esimerkiksi taulukkomuotoon.
4. Käytä tilastotestejä tulkitsemaan aineiston ryhmien keskinäisiä suhteita.
5. Tulkitse tulokset vahvasti teoriaan nojaten.

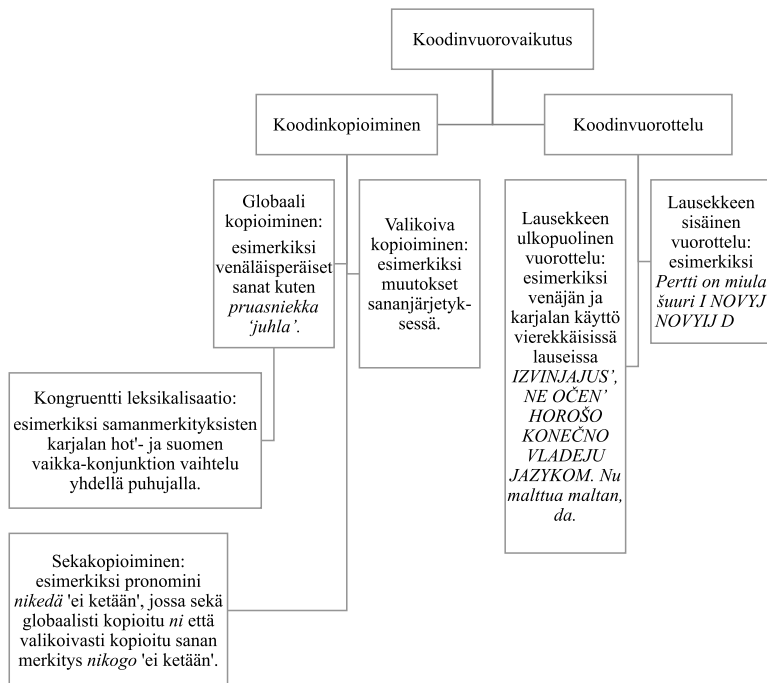
Menetelmä ei ole uusi, vaan sitä on hyödynnetty korpuslingvistiikassa runsaastikin. Menetelmä tukee kontaktilingvistiikan käyttöpohjaisen näkökulman (esim. Backus 2014) toteutumista, ja tutkimus on luonteeltaan korpus- eli aineistolähtöistä (McEnery, Xiao & Tono 2006: 8). Aineistolähtöisessä tutkimuksessa valittuun teoreettiseen viitekehykseen nojaaminen on tärkeää, koska teorian valossa tulokset voidaan selittää kattavasti (ks. esim. Gries & Stefanowitsch 2006: 73). Näin korpuslingvistinen metodologia tarjoaa vahvan aineistolähtöisen pohjan tutkimukselle.

Toinen tärkeä kohtaamani ongelma liittyy kontaktilingvistisiin teorioihin ja niiden termien selitysvormaisuuteen. Kaksikielisyyttä sekä muunkinlaista kielten välistä vaikutusta on kuvattu monin eri kontaktilingvistisin teorioin. Yhtä yhtenäistä mallia kuvaamaan kontaktilingvistisiä ilmiöitä ei kuitenkaan ole löydetty. Erityistä kritiikkiä teorian ovat saaneet termien merkityksiin liittyvistä ongelmista sekä erilaisista tyypologisista rajoituksista, jotka eivät päde kaikissa tilanteissa (ks. esim. Verschik 2008: 54). Teorioita yhdistää jako ajallisesti yksilotteiseen eli synkroniseen näkökulmaan ja kielenkäytön kehittymistä kuvaavaan eli diakroniseen näkökulmaan. Karkeammin tämä asia voidaan esittää esimerkin avulla: monet sanat on saatu kieliin jo vuosisatoja tai -tuhansia sitten. Tällaisia ovat suomen ja karjalan venäläisperäiset lainasana-esimerkkiklassikot *pappi*, *risti* ja *päčči* 'uuni'. Nämä ovat esimerkki diakronisesta näkökulmasta. Esimerkkejä synkronisesta näkökulmasta ovat edellä esittämäni karjalan kielen synonyymit eli kontaktilähtöinen variaatio. Lisäksi esimerkkinä on sanomalehtiaineistossani käytetty suomen kielen sana ”*pipo*” lainausmerkeissä osoittamassa sanan vakiintumattomuutta.

Tutkimuksessani päädyin hyödyntämään koodinkopioimisen viitekehystä (*the code-copying framework*), jonka on kehittänyt Lars Johanson turkkilaisten kielten kontaktien pohjalta (esim. 1999, 2002). Teoriaa on myös kehittänyt Anna Verschik virolaisvenäläisen kaksikielisyyden pohjalta (2008). Teorian suurimpia etuja on runsaankin

kopioimisen hyväksyminen ilman, että sen ajatellaan koituvan kielen kohtaloksi. Kielen rakenteelliset seikat siis erotetaan sosiaalisista ulottuvuuksista. Lisäksi teoria on lingvivistisestä näkökulmasta salliva: sen puitteissa voidaan tarkastella diakronista ja synkronista näkökulmaa yhdessä eikä typologisia rajoituksia kopioimiselle aseteta.

Koodinkopioimisen viitekehyksessä kaksikielinen kielenkäyttö nähdään koodinvuorovaikutuksena (*code interaction*), jossa kielistä käytetään termiä koodi. Kieli, jossa on toisen kielen vaikutusta, on peruskoodi (*basic code*), ja vaikutusta antavaa kieltä kutsutaan mallikoodiksi (*model code*). Koodinvuorovaikutus jakautuu edelleen koodinkopioimiseen (*code copying*) ja koodinvuorotteluun (*code alternation*). Koodinkopioiminen jakautuu globaaliin (*global copying*), valikoivaan (*selective copying*) ja sekakopiointiin (*mixed copying*). Ensimmäinen vastaa lausuman sisäisiä perinteisiä lainasanoja sekä koodinvaihtoa, toinen rakenteellista lainaamista ja esimerkiksi käännöslainoja ja kolmas on kahden ensimmäisen sekamuoto. Globaalia kopioimista olen täydentänyt Verschikin (2008: 70) mukaisesti Peter Muyskenin (2000) termillä kongruentti leksikalisaatio (*congruent lexicalization*), joka kuvailee kahden läheisen sukukielen leksikaalisten aineiden paradigmaattista vaihtelua. Koodinvuorottelussa erotellaan lausuman sisäiset (*intraclausal*) ja ulkoiset (*extraclausal*) sekvenssit. (Termien kuvauksesta ks. Johanson 2002: 291–292; Verschik 2008: 54, 61–62.) Kuvio 1 havainnollistaa viitekehysten termien keskinäisiä suhteita karjalaisessa kontekstissa esimerkein.



**Kuvio 1.**  
Koodinvuorovaikutus karjalaisessa kontekstissa

Koodinkopioimisen viitekehys keskittyy lausuman sisäisiin (*insertional*) ilmiöihin, joten koodinvuorottelu jää vähäisemmälle osalle. Tutkimuksessani olen kuitenkin määritellyt myös koodinvuorottelua, koska se on osa kaksikielisiä resursseja karjalaisessa kielenkäytössä. Koodinkopioimisen viitekehyksessä koodinvuorottelu nähdäänkin osana keskustelua eikä niinkään itse kieltä eli koodia. Se auttaa näkemään sekoittumisen osana keskustelua. Näin ollen ei siis ajatella, että kieli olisi sekakieli, kuten karjalan yhteydessä usein on hieman negatiiviseen sävyyn puhuttu, vaan ajatellaan, että sekoittuminen johtaa sekadiskurssiin. Tämä on etenkin vähemmistökielen elvyttämisen kannalta tärkeä ideologia, koska kaksikieliset resurssit ovat osa vuorovaikutuksen strategioita.

Väitöstutkimukseni artikkeleissa ilmeni, että kaksikielisiä resursseja eli koodinkopioimista ja koodinvuorottelua käytetään pragmaattisista syistä. Esimerkiksi koska karjalaa puhutaan runsaasti Venäjällä ja puhujat ovat saaneet koulutuksensa vain venäjäksi, suurimmat numeraalit sekä esimerkiksi ammatteja ja yhteiskuntaa koskevat asiat tuotetaan venäläisin resurssein. Omaperäisiä elementtejä ovat erityisesti konkreettiset perussanat ja esimerkiksi kokoa ja värejä kuvaavat adjektiivit. Suomalaisperäistä taas on partikkeleiden kaksikielinen käyttö sekä tietoiset vaihtoehdot venäläisperäiselle sanastolle: tätä ilmenee niin puhutussa suomenkarjalassa kuin molempien valtioiden alueella tapahtuvassa kielen tietoisessa kehittämisessä. Koodinvuorottelua taas ilmenee useammin venäjänkarjalaisessa diskurssissa. Väitöstutkimukseni tulokset vahvistavatkin Peter Auerin (2007) kohta jo parin vuosikymmenen takaista huomiota: kaksikielinen puhe ei ole kahden yksikielisen puhujan summa vaan aivan omanlainen kokonaisuutensa, jossa resurssien käyttö muodostaa oman hienojakoisen säännöstönsä.

Lopuksi on todettava, että kontaktilingvistiksi teoriassa on vielä runsaasti töitä tehtävänä: yleensä kaksikielistä kielenkäyttöä koskevat teoriat ja termit eivät ole veyneet kuvaamaan aivan kaikkia kielenkäytön ilmiöitä tai ainakaan niitä ei ole saatu koskemaan kaikkia kaksikielisiä. Muistankin törmänneeni lauseeseen ”Onko kaksikielisessä puheessa muuta universaalia kuin yksikielisyyden puuttuminen?” Työsarkaa on myös karjalan kuvaamistavassa, joka arvostaisi kieltä ja sen käyttäjiä parhaalla mahdollisella tavalla, sekä tutkimusmenetelmien tehostamisessa, kun resurssit pienen kielen kohdalla ovat luonnollisesti pienet. Toivon kuitenkin, että tämä väitöskirjani on ottanut harppauksen kohti näiden tavoitteiden toteutumista.

## Lähteet

- AUER, PETER 2007: The monolingual bias in bilingualism research, or: why bilingual talk is (still) a challenge for linguistics. – Monica Heller (toim.), *Bilingualism. A social approach* s. 319–339. London: Palgrave Macmillan.
- BACKUS, AD 2014: A usage-based approach to borrowability. – Zenner & Gitte Kristiansen (toim.), *New perspectives on lexical borrowing. Onomasiological, methodological and phraseological innovations* s. 19–39. Boston: Mouton De Gruyter.
- GILOEVA, NATALIA 2015: Yle uudizet karjalakse: ylen hyvä! <http://kianna-hanke.blogspot.com/2015/04/yleuudizet-karjalakse-ylen-hyva.html> (7.4.2015).
- GRIES, STEFAN T. – STEFANOWITSCH, ANATOL 2006: Cluster analysis and the identification

- of collexeme classes. – Sally Rice & John Newman (toim.), *Empirical and experimental. Methods in cognitive/functional research* s. 73–90. Stanford, CA: CSLI Publications.
- HAUGEN, EINAR 1950: The analysis of linguistic borrowing. – *Language* 20 s. 210–223.
- JOHANSON, LARS 1999: Dynamics of code-copying in language encounters. – Bernt Brendemoen, Elizabeth Lanza & Else Ryen (toim.), *Language encounters across time and space* s. 37–62. Oslo: Novus Press.
- 2002: Contact-induced change in a codecopying framework. – Mari C. Jones & Edith Esch (toim.), *Language change. The interplay of internal, external and extralinguistic factors* s. 286–313. Berlin: Mouton de Gruyter.
- KOVÁCS, MAGDOLNA 2009: Koodinvaihto ja kielioppi. – Jyrki Kalliokoski, Lari Kotilainen & Päivi Pahta (toim.), *Kielet kohtaavat* s. 24–49. Tietolipas 227. Helsinki: Suomalaisen Kirjallisuuden Seura.
- KUUSI, PÄIVI – RIIONHEIMO, HELKA – KOLEHMAINEN, LEENA 2022: Translating into an endangered language. Filling in lexical gaps as language making. – *International Journal of the Sociology of Language* 274 s. 133–160. <https://doi.org/10.1515/ijsl-2021-0019>.
- MCENERY, TONY – XIAO, RICHARD – TONO, YUKIO 2006: *Corpus-based language studies. An advanced resource book*. London: Routledge.
- MUYSKEN, PIETER 2000: *Bilingual speech. A typology of code-mixing*. Cambridge: Cambridge University Press.
- SARHIMAA, ANNELI 1999: *Syntactic transfer, contact-induced change, and the evolution of bilingual mixed codes*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- TAVI, SUSANNA 2018: Rajakarjalaismurteiden kielikontaktit venäläiskopioiden taajuuden ja fonologian valossa. – *Lähivörtlusi. Lähivertailuja* 28 s. 316–356. <https://doi.org/10.5128/LV28.10>.
- 2022: Lexical outcomes of Karelian-Russian bilingualism in Tver Karelian. – *Eesti ja soome-ugri keeleteaduse ajakiri* 13 (2) s. 347–381. <https://doi.org/10.12697/jeful.2022.13.2.11>.
- TAVI, SUSANNA – TAVI, LAURI 2019: 'Ja', 'tai'- ja 'vaikka'-merkityksisten lekseemien kontaktiähtöinen variaatio rajakarjalaismurteissa. – *Virittäjä* 123 (3) s. 401–429. <https://doi.org/10.23982/vir.70346>.
- 2022: Vocabulary-based classification and contact-induced formation of neologisms in two standard varieties of Karelian. – *Journal of Language Contact* 14 (3) s. 684–719. <https://doi.org/10.1163/19552629-14030006>.
- WEINREICH, URIEL 1968 [1953]: *Languages in contact. Findings and problems*. The Hague: Mouton & Co.
- VERSCHIK, ANNA 2008: *Emerging bilingual speech. From monolingualism to code-copying*. London: Continuum International Publishing Company.

**Susanna Tavi:** *Contact-induced lexical variation in the Karelian language. Corpus-based perspectives on code-copying*. Publications of the University of Eastern Finland Dissertations in Education, Humanities and Theology No 197. Joensuu: University of Eastern Finland 2023. <http://urn.fi/URN:ISBN:978-952-61-4779-6>.

Kirjoittajan yhteystiedot:  
etunimi.sukunimi@uef.fi