

Karjalan Sanomat -korpus Petroskoin (käännös)suomen piirteitä

Jukka Mäkisalo & Hannu Kemppanen
Itä-Suomen yliopisto

Anna Saikonen
Petroskoin valtiollinen yliopisto

Abstract

The present article introduces a minority language corpus, the Newspaper Corpus of Karelian Finnish, and the tentative results of lexical and grammatical analysis based on it. The corpus was compiled from the newspaper *Karjalan Sanomat* and consists of two sub-corpora: texts originally written in Finnish and texts translated into Finnish from Russian. This is the first attempt to analyse a minority language variant of Finnish spoken outside Finland by comparing it to Standard Written Finnish (SWF) and using corpus linguistic methods. The two sub-corpora of *Karjalan Sanomat* are compared to the newspaper *Karjalainen* published in Eastern Finland, and the objective is to gain a lexical and grammatical cross-section based on corpus linguistic methods. The analysed features include some quantitative lexical indicators (word length, sTTR), distribution of syntactic categories, and key words (lexical contents and morphosyntactic categories). The comparison shows that the two variants of Karelian Finnish resemble each other more than SWF. Furthermore, it is theoretically interesting that a translated minority language variety is lexically richer (less repetitive) than originally written minority language variety, a result opposite to earlier comparisons between translated and non-translated language variants within majority languages.

Keywords: Finnish language, corpus, minority language, translated language

Avainsanat: suomen kieli, korpus, vähemmistökieli, käännöskieli

1 Johdanto

Tässä artikkelissa esiteltävä korpusaineisto ja analyysi avaavat uuden näkökulman käännetyn kielen tutkimukseen. Tarkastelun kohteena on Itä-Suomen yliopiston ja Petroskoin yliopiston yhteistyönä kokoama Karjalan suomen sanomalehtikorpus. Käännöstieteen näkökulmasta aineiston tekee kiinnostavaksi se seikka, että osa sanomalehden materiaaleista kirjoitetaan suoraan suomeksi, osa puolestaan tuotetaan kääntämällä tekstit venäjältä. Uuden näkökulman käännetyn kielen tutkimukseen tarjoaa

analyysi, jossa vertaillaan suomen eri kielimuotoja, jotka edustavat erilaisia kielen tuottamisen olosuhteita akselilla vähemmistökieli–enemmistökieli. *Karjalan Sanomat* -sanomalehtikorpusista verrataan Pohjois-Karjalassa ilmestyvään *Karjalainen*-sanomalehden korpusaineistoon, joka edustaa Suomessa tuotettua vastaavaa aineistoa.

Aiempaan käännetyn kielen tutkimukseen suhteutettuna nyt esiteltävä tutkimuskohde edustaa käännettä tutkimuksessa toistaiseksi vähän tarkasteltua aluetta – vähemmistökielille (käsitteenä ks. alla luvussa 4) kääntämistä. Vähemmistökielen puhujien viestintäympäristö muodostuu jatkuvista kielikontaktitilanteista, jotka edellyttävät joskus kääntämistä sanan varsinaisessa merkityksessä, joskus taas koodin vaihtoa kielen vaihtuessa toiseen viestintätilanteen mukaisesti. Vähemmistökielen käyttäjien omat tuotokset altistuvat jatkuvasti enemmistökielen vaikutukselle. ”Kääntäminen” on näin ollen kognitiivisesti vahvasti läsnä myös sellaisessa kielen tuottamisessa, jota ei ole perinteisesti tarkasteltu käännettä tutkimuksessa. Kielikontaktit ovatkin kuuluneet selvästi lingvistiseen tutkimuskenttään. Kääntämisen läsnäolo kielikontakteissa on jäänyt toistaiseksi vähälle huomiolle myös lingvistisessä lähestymistavassa. Vasta aivan hiljattain on huomattu, että kielikontakteilla ja kääntämisellä sekä lisäksi toisen kielen omaksumisella on ilmiönä toisiaan muistuttavia ja toisiinsa kytkeytyviä piirteitä. Suomessa *Virittäjän* erikoisnumero vuodelta 2014 nostaa ensimmäistä kertaa esille näiden tutkimuksen alueiden yhtymäkohdat (Riionheimo ym. 2014).

Karjalan Sanomat -korpuksen rakenne antaa mahdollisuuden vertailla kieliaineistoja, jotka julkaistaan muodollisesti yhtenäisenä aineistona yhden ja saman median sivuilla, mutta jotka kuitenkin edustavat tuotantotavaltaan erilaisia kielimuotoja. Lisäksi molemmat tarkasteltavat kielimuodot – yhtäältä alun perin suomeksi, ilman käännettäprosessia ja toisaalta nimenomaan kääntäen laaditut sanomalehtitekstit – on tuotettu Suomessa laadituista lehtiteksteistä poikkeavassa viestintäympäristössä. Tekstejä ovat tuottaneet kirjoittajat, joille suomi on toinen päivittäisessä viestinnässä käytettävistä kielistä, mutta sen käyttö voi olla venäjänkielisessä ympäristössä hyvinkin rajoittunutta eikä todennäköisesti kata niin laajaa kielenkäyttötilanteiden kirjoa kuin yksikielisessä ympäristössä suomea käytettäessä. On myös otettava huomioon, että suomen kieltä tuotetaan Suomen ulkopuolella, eristyksissä suomalaisen kieliyhteisön kielinormeista. Venäjän Karjalassa paikallinen kielenhuolto koostuu harvalukuisesta kielentaitajien joukosta, ja on mahdollista, että Suomessa sovellettava normisto ei ulotu kaikessa voimassaan Venäjän suomenkielisen väestön käyttämään kieleen, vaan syntyy omia normeja.

Pyrimme tässä artikkelissa luomaan läpileikkauksen tarkasteltaviin kielimuotoihin keskittymällä muutamaaan sanastolliseen ja morfosyntaktiseen piirteeseen. Vertaamalla Karjalan suomen molempien kielimuotojen ominaisuuksia Suomessa tuotettuun ”enemmistösuomeen” voidaan saada tietoa siitä, mitä eroja näiden kolmen eri kielimuodon välillä on, jos on. Lisäksi saadaan tietoa siitä, onko vähemmistökielestä kääntämisellä joitakin ominaisuuksia, jotka tuovat uutta tietoa väittämiin käännetyn kielen erityispiirteistä.

2 Karjalan Sanomat -korpus

Karjalan Sanomat -korpus on Venäjän Karjalan suomen kielen sanomalehtikorpus. Se on koottu Itä-Suomen yliopistossa yhteistyössä Petroskoin valtionyliopiston kanssa SILK-projektin (Suomen itäisten lähialueiden kielikorpuksien) puitteissa, jonka tavoitteena on ollut (1) luoda rajakarjalaismurteiden, (2) inkeriläismurteiden sekä (3) Venäjän Karjalan yleiskielen korpuksia. Hanke on saanut rahoitusta Koneen säätiöltä vuosina 2012–2015.

Kaikkiaan korpuksen valinnan ja kokoamisen kriteereinä on käytetty samoja periaatteita kuin vertailuaineistona käytettävän *Karjalainen*-sanomalehden korpuksen¹ kokoamisessa. Sanomalehtikorpuksen kokoaminen aloitettiin vuonna 2013 Itä-Suomen yliopistossa. Tekstiaineistoa koottiin ja korpusta tehtiin syyskuusta 2013 huhtikuuhun 2014. Aineistoksi valittiin Petroskoissa ilmestyvän *Karjalan Sanomat* -lehden 2000-luvun tekstejä, ja se on koottu seuraavien valintakriteerien mukaan: (1) tekstit edustavat yleiskieltä, ja (2) ne ovat toimituksen tuottamia kokonaisia lehtitekstejä, mikä tarkoittaa että erilaiset ilmoitukset, mainokset yms. on jätetty pois. Korpus on koottu lehden painetusta versiosta eli toimituksesta suoraan saaduista, lehden painoon menevistä tiedostoista, koska ne ovat artikkeleina täydellisempiä.²

Karjalan Sanomat on ainoa Venäjän Karjalassa ilmestyvä suomenkielinen viikkolehti. Lehdessä kerrotaan muun muassa taloudesta, kulttuurista, urheilusta, koulutuksesta, Karjalan eri (alueellisten) piirien elämästä sekä tasavallan yhteiskuntapoliittisista tapahtumista. Erityistä huomiota lehti kiinnittää nuoriin lukijoihin, joille julkaistaan omia palstoja: *Lasten maailma*, *Nuorison ääni* ja *Suomen kieli*. Lehden sivuilla käsitellään myös kansallisuuspolitiikkaa ja kielikysymyksiä, jotka ovat olleet Karjalassa kipupisteitä muutaman viime vuosikymmenen aikana. ”[S]uomesta on yhä enenevässä määrin tulossa Karjalassa vieras kieli, mikä ilmenee muun muassa suomenkielisen taidekirjoittamisen laskuna. Syynä on karjalan kielen voimakas pyrkimys erottautua lähisukukielestään suomesta [- -]” (Sundelin 2000: 61). Karjalassa puhuttavan suomen kielen suosion vähenemisen syinä voidaan myös objektiivisesti pitää seuraavia tekijöitä: (1) monien lehden lukijoiden ja toimittajien muuttamista Suomeen, (2) siirtymistä sosialistisesta järjestelmästä markkinatalouteen ja (3) paikallisen hallituksen harjoittamaa karjala- ja vepsämielistä politiikkaa.

Korpuksen kokonaissanamäärä on 582 166 sanetta, ja se jakautuu kolmeen osaan: alun perin suomeksi kirjoitetut tekstit (75,78 %), venäjältä käännetyt tekstit (10,87 %) ja kahteen edelliseen ryhmään määrittelemättömiksi jääneet tekstit (13,35 %). Korpus koostuu 2169 tekstitiedostosta, ja jokainen tiedosto vastaa yhtä kokonaista lehtiartikkelia. Joissakin tiedostoissa voi olla kaksi tai jopa kolme artikkelia, joista ensimmäinen on aina pääuutinen ja jälkimmäiset ovat ns. kainalojuttuja eli pääuutisen aiheeseen liittyviä ja sisällöltään sitä täydentäviä kommenttimuotoisia tekstejä. Jokaisen tiedoston sisälle on tallennettu myös jutun otsikko, leipäteksti ja kirjoittajan nimi (joidenkin pienten uutisten kirjoittajia lehdessä ei mainita). Korpuksen on otettu vuoden 2012 jälkipuoliskon, koko vuoden 2013 ja vuoden 2014 alkupuoliskon lehtien numerot. Lehtiaineiston tärkeänä

¹ Noin 40-miljoonainen korpus koottiin Joensuun yliopistossa noin 15 vuotta sitten.

² 2000-luvun alussa lehti siirtyi Internetiin. Vuodesta 2012 lukijat ovat voineet tilata lehden sähköpostiinsa pdf-muotoisena.

valintakriteerinä on ollut se, että korpukseen on hyväksytty vain kokonaiset uutiset ja lehtiartikkelit; se ei sisällä horoskooppeja, mainoksia, ilmoituksia, kuvatekstejä eikä yle.fi-sivuilta lainattuja uutisia. – Lähitulevaisuudessa korpus julkaistaan Kielipankissa.

3 Käännetty kieli korpustutkimuksen kohteena

Käännetyin kielen piirteitä on tarkasteltu korpuslingvistisin menetelmin ensisijaisesti analysoimalla suuria, laajalle levinneitä kieliä. Englannin kielen piirteet on oletettavasti kartoitettu parhaiten. Tähän on myös historialliset syynsä: korpuslingvistinen käännöstutkimus ja ns. käännösuniversaalien tutkimus saivat alkunsa Manchesterin UMIST-yliopistossa. Käännetyllä kielellä oletettiin olevan tiettyjä, kieliparista riippumattomia piirteitä, jotka erottavat käännökset ilman käännösprosessia tuotetuista, ei-käännetyistä teksteistä. (Baker 1993; Laviosa-Braithwaite 1996.) Suomen kielen osalta universaalien tutkimus pääsi alkuun vain vähän myöhemmin (ks. esim. Mauranen 2000; Tirkkonen-Condit 2002; Eskola 2002; Mauranen & Kujamäki 2004).

Käännösten kielellisistä (ensisijaisesti leksikaalista) ominaisuuksista on tehty useita eri hypoteeseja. Kiinnostuksen kohteena ovat olleet mm. seuraavat ilmiöt: yksinkertaistuminen, normalisoituminen, standardisoituminen, eksplisiittistuminen, ns. *levelling out* eli pyrkimys valita käännöksiin tarjolla olevien leksikaalisten yksiköiden jatkumolta sen keskelle sijoitettavia elementtejä, uniikkiainesten kääntäminen, epätyypilliset kollokaatiot sekä interferenssi. (Ks. yleiskatsaus universaalien tutkimuksesta Zanettin 2012 ja Kuusi 2011.)

Korpuslingvistisen käännöstutkimuksen ensimmäinen sukupolvi keskittyi pääsääntöisesti universaalihypoteesien testaamiseen. Vaikka korpustutkimusta on hyödynnetty sitten käännöstutkimuksessa laajemminkin, on Zanettinin (2012: 23) mukaan kääntämisen universaaleilla ollut aivan viime vuosiin saakka edelleen keskeinen asema korpuslingvistisessä käännöstutkimuksessa. Muutamat viime aikoina ilmestyneet tutkimukset ovat kuitenkin osoittaneet, että universaalien asema korpustutkimuksen keskiössä on alkanut muuttua, ja on ryhdytty kyseenalaistamaan universaalien olemassaoloa. Gert De Sutter ym. (2012: 138) nostavat esille kaksi keskeistä seikkaa, jotka asettavat klassiset universaalihypoteesit kyseenalaisiksi. Ensinnäkin monet perinteisistä hypoteeseista ovat motivoimattomia, monimutkaisia ja epäselvästi muotoiltuja. Toiseksi korpuspohjainen käännöstiede ei ole ottanut huomioon lähdekielen ja tekstilajin mahdollista vaikutusta. Tällaisten väitteiden esittäminen nostaa universaalit jälleen keskusteluun, mutta nyt aiempiin tutkimuksiin kohdistuvana kritiikkinä. Keskustelun herättäneistä käännöstieteilijöistä mainittakoon esimerkiksi Juliane House (2008), Viktor Becher (2010) sekä Silvia Bernardini ja Adriano Ferraresi (2011).

4 Vähemmistökielet ja kääntäminen: korpuslingvistinen lähestymistapa

Tässä artikkelissa tarkasteltava tutkimuskysymys on osa vähemmistökielten ja kääntämisen problematiikkaa. Kuten Albert Branchadell (2011: 98) toteaa, aihepiiri ei ole ollut käännöstieteen keskiössä viimeisen 25 vuoden aikana, vaikka jo Gideon Toury

(1985: 7) peräänkuulutti kääntämisen tärkeää roolia vähemmistökielten säilymisessä ja kehittämisessä niiden yrittäessä olla joutumatta enemmistökielten korvaamiksi.

Vähemmistökielten ja kääntämisen problematiikkaa on tarkasteltu ensi sijaisesti tutkimuksissa, jotka edustavat ns. kulttuurisen käänteeseen myötä syntyneitä lähestymistapoja. Lawrence Venuti (1998: 135) tuo vähemmistöille ja kääntämiselle omistetun *Translator*-julkaisun johdannossa selkeästi esille kriittisen tarkastelukulman. Hänen mukaansa 'vähemmistö' tarkoittaa alisteista kulttuurista tai poliittista asemaa. Sosiaalinen konteksti määrittää vähemmistön paikalliseksi, kansalliseksi tai globaaliksi. Tämän aseman saavat ne kielet ja kirjallisuudet, joilta puuttuu arvovalta ja auktoriteetti. Ne ovat standardisoimattomia ja kanonisoimattomia, valtakulttuurit eivät puhu tai lue niitä. Venuti korostaa, että 'enemmistö' ja 'vähemmistö' ovat suhteellisia käsitteitä ja riippuvaisia toisistaan. Lisäksi ne ovat aina riippuvaisia historiallisesta tilanteesta, myös silloin kun tilanne muuttuu.

Vähemmistökielen käsite liitetään tässä yhteydessä sekä aikaan että paikkaan. Kieli voi olla tiettyä aikana jossakin yhteiskunnassa enemmistökielenä ja muuttua myöhemmin vähemmistökieleksi tai päinvastoin. Esimerkiksi englannin kieli oli vuosisatoja Irlannissa vähemmistökielenä, kunnes se muuttui poliittisten toimien, nälänhädän ja iirinkielisten valtavan maastamuuton seurauksena enemmistökieleksi 1800-luvulla. Diakroniset muutokset kytkeytyvät usein spatiaalisiin muutoksiin. Esimerkiksi venäjän kielestä on tullut Baltian maissa vähemmistökieli Neuvostoliiton hajottua. (Cronin 1995: 86–87.)

Vähemmistökielen problematiikka sisältää useita kysymyksiä, joita voidaan tarkastella käännöstutkimuksessa. Michael Cronin (1995: 87–89) tuo esille kolme keskeistä seikkaa. Ensinnäkin kielen enemmistöasema syntyy poliittisten, taloudellisten ja kulttuuristen muutosten seurauksena. Näin ollen kaikki kielet ovat potentiaalisia vähemmistökieliä. Toiseksi kielten välillä voidaan nähdä eroja sen mukaan, kuinka paljon kuhunkin kieleen tai kustakin kielestä käännetään. Osa kielistä on kohdekieli-intensiivisiä, osa lähdekieli-intensiivisiä. Englanti on hyvä esimerkki kohdekieli-intensiivisestä kielestä eli kielestä, josta siirtyy paljon tekstejä käännöksinä kohdekieliin. Suomi käy puolestaan esimerkiksi lähdekieli-intensiivisestä kielestä, ts. muut kielet ruokkivat suomalaista kulttuuria lukuisilla käännöksillä. Kolmantena tekijänä Cronin ottaa esille kääntämisen keskeisen merkityksen vähemmistökielille. Kääntäminen toimii kyseisen kielen olemassaolon jatkeena ja itseluottamuksen vahvistajana.

Kääntämisen rooli vähemmistökielen säilymisessä on alettu ymmärtää vasta viime aikoina. Myös tähän aihepiiriin liittyvää tutkimusta on ryhdytty tekemään. Suomalais-ugrilaisten kielten osalta kääntämisen roolia vähemmistökielen säilymisessä tutkitaan Itä-Suomen yliopistossa vuosina 2015–2018 toteutettavissa *Kiännä-* ja *Kontu-*hankkeissa. Tutkimuksen lisäksi hankkeisiin liittyy myös konkreettisia kielen (tässä tapauksessa karjalan kielen) elvyttämiseen tähtäviä toimenpiteitä, kuten kursseja karjalan kielen kääntämisestä kiinnostuneille.

Vähemmistökieliä ja kääntämistä on tarkasteltu lähes yksinomaan kulttuurintutkimuksen ja sosiologisesti suuntautuneen käännöstutkimuksen piirissä. Empiirinen tarkastelu on toistaiseksi jäänyt vähälle. Kulttuuritutkimuksellista otetta on tosin konkretisoitu empirialla Anthony Pymin ja Grzegorz Chrupałaan (2005)

tutkimuksessa, jossa he tilastoivat käännösvirtoja pienten ja suurten kielten välillä ja vertasivat käännösten vs. alkukielisten teosten lukemista eri kulttuureissa.

Vähemmistökielet ja kääntäminen ei tietävästi ole vielä ollut korpuslingvistisen käännöstutkimuksen kohteena, vaikka kieli- ja käännösteknologia onkin nähty hyödyllisenä apuvälineenä vähemmistökielten tallentamisessa ja elvytysprosessissa. Esimerkiksi Etelä-Afrikassa, jossa on 11 virallista kieltä, on koottu paralleelikorpuksia auttamaan kääntäjiä erityisesti terminologisessa työssä (Madiba 2004). Samoin konekääntämisessä on Branchadellin (2011: 99) mukaan panostettu viime vuosina paljon vähemmistökielten kääntämiseen. Huomionarvoisia ovat mm. kansainväliset SALTMIL-työpajat (Speech and Language Technology for Minority Languages), joissa kiinnostuksen kohteena on ollut kieli- ja käännösteknologian hyödyntäminen vähemmistökielten tutkimukseen sekä kielten säilymistä ja elvyttämistä edistäviin käytännön toimiin.

Tässä artikkelissa esiteltävä korpusaineisto ja analyysi avaavat tietyn uuden näkökulman käännetyin kielen tutkimukseen. Itse korpusaineiston kokoaminen antaa mahdollisuuden tallentaa edustava näyte kahdesta Venäjän Karjalassa käytetystä kielimuodosta. Tarkastelun kohteeksi otetaan vähemmistökieleen tehdyt käännökset, joiden kielellisiä piirteitä suhteutetaan ilman käännösprosessia vähemmistökielillä tuotettujen tekstien piirteisiin ja vastaavaan enemmistökieliseen muotoon. On oletettavaa, että analyysi antaa uutta tietoa käännösten kielellisiin piirteisiin vaikuttavista tekijöistä.

5 Tutkimusmenetelmät

Tässä tutkimuksessa on tarkasteltu kahden kielimuodon välillä sellaisia leksikaalisia (ja leksikologisia) sekä syntaktisia piirteitä, joiden esiin nostamisessa voidaan hyödyntää korpusmenetelmiä. Ohjelmanä on käytetty *WordSmith Toolsin* (WST) versiota 5.0 ((c) Mike Scott).

Karjalan Sanomien kahta kielimuotoa, alun perin suomeksi kirjoitettua ja venäjältä käännettyä suomea, on verrattu vastaavaan Suomessa kirjoitettuun lehtitekstiin, nimittäin *Karjalainen*-sanomalehden vastaavanlaisin kriteerein koottuun korpukseseen. Lisäksi, kun tarkastelemme sanaston kvantitatiivisia piirteitä, käytämme osittain vertailukorpuksina Käännössuomen korpuksen (Mauranen 2004) viihde- ja tietokirjallisuuden sekä supi- että käännössuomen (ss ja ks) osakorpuksia ja Yle-korpuksen ruututekstien korpusta (ks. sanamäärät taulukosta 1).

Taulukko 1: Osakorpuksen sanamäärät

| | supisuomi | käännössuomi |
|--------------------|-----------|--------------|
| viihdekirjallisuus | 195 602 | 206 859 |
| tietokirjallisuus | 506 665 | 534 060 |
| ruututekstit | - | 1 000 000 |
| Karjalainen | 1 003 881 | - |
| Karjalan Sanomat | 441 182 | 63 292 |

Vertailun kohteeksi on valittu seuraavat kielen piirteet:

- (1) sanaston kvantitatiivisista ominaisuuksista sanapituus sekä standardoitu type/token-suhdeluku (sTTR)
- (2) sanojen syntaktisten kategorioiden (sanaluokkien) pääluokkien jakauma, ts. nominien, verbien ja funktiosanojen suhteellinen osuus
- (3) korpusten frekventeimmän sanaston sisältö ja muotoryhmät.

Sanapituus korreloi yleensä selvästi tekstin vaikealukuisuuden ja abstraktiuden kanssa niin, että mitä pitemmät sanat, sitä vaikealukuisemmaksi ja abstraktimmaksi teksti koetaan (Mikk 2008). sTTR taas heijastaa tekstin toisteisuutta sanatasolla: mitä korkeamman arvon 1–100 välillä teksti saa, sitä vähemmän tekstissä on toisteisuutta eli sitä rikkaampaa tekstin sanasto. Frekventeimmän sanaston sisältöä tarkastellaan WST:n sanalista-toiminnon avulla ja muotoryhmiä avainsana-toiminnon avulla. Jälkimmäisellä analyysillä tarkastellaan käytettyjen kieliopillisten kategorioiden eroja.

Avainsana-analyysia on käytetty käännöstieteellisessä korpustutkimuksessa tarkasteltaessa verrannollisten tekstien leksikaalisia piirteitä. Verrannollisuus tarkoittaa tässä yhteydessä käännettyjen ja ei-käännettyjen tekstien piirteitä. Avainsana-analyysin kohteena ovat olleet muun muassa kääntäjän idiosynkraattiset tyylipiirteet (Kenny 2000) ja puhekielen piirteet käännetyissä kaunokirjallisuudessa (Nevalainen 2004), tietokirjallisten tekstien diskursiiviset piirteet (Kemppanen 2004, 2008) sekä käännettyjen sanomalehtitekstien ajankohtaisia aiheita kielentävä sanasto (Laviosa 2002). Avainsana-analyysia on käytetty myös käännöstutkimuksen kannalta relevantissa lingvistisessä korpustutkimuksessa, kuten tyyliintutkimuksessa (Culpeper 2012; Mahlberg 2013), diskurssianalyysissa (Baker 2006) ja oppijankielen tutkimuksessa (Jantunen 2011).

Tämän tutkimuksen kiinnostuksen kohteena olevasta aineistosta voidaan avainsana-analyysin avulla nostaa esiin sanaston sisältöä peilaavia leksikaalisia elementtejä sekä morfosyntaktisia muotoryhmiä, jotka antavat tietoa kielenkäytön erityispiirteistä. Kun kyseessä ovat vähemmistökielenä tuotetun suomen kielen kaksi varianttia – alun perin suomeksi kirjoitetut sanomalehtitekstit ja samaa genreä edustavat venäjämäiset käännettyt tekstit – avainsana-analyysi voi tehdä näkyväksi niiden kielimuotojen erot.

Tämän tutkimuksen näkökulmasta relevanttia avainsana-analyysia edustaa lähestymistapa, jossa tuotettua avainsanalistaa käytetään tekstin keskeisen sisällön tarkasteluun. Korpustutkijat käyttävät alun perin Martin Philipsin (1989) luomaa termiä *aboutness* viittaamaan tekstin keskeiseen sisältöön (Scott 2000; Baker 2006; Culpeper 2012): avainsanat sisältävät muun muassa tiivistettynä tiedon siitä, mistä teksti kertoo. Tiivis, tekstin sisältöön viittaava tieto on käyttökelpoinen korpustyökalu erityisesti tyylin ja diskurssintutkijoiden käyttöön. Käännöstutkimuksessa sillä voidaan saada tietoa käännetyn ja ei-käännetyn kielimuodon mahdollisista eroista.

Korpusohjelmalla tuotettu avainsanalista sisältää erityyppisiä leksikaalisia elementtejä. Avainsanoja on jaettu tutkimusintressien mukaisesti eri ryhmiin. Tyypillinen luokittelu on jako kolmeen ryhmään: 1) erisnimet, 2) tekstin sisältöön viittaavat sanat ja

3) tekstin tyyliä implikoivat sanat. (Scott & Tribble 2006; Mahlberg 2013: 67.) Tässä analyysissä kolmatta ryhmää käsitteellistetään lähinnä kieliopillisten funktioiden kautta.

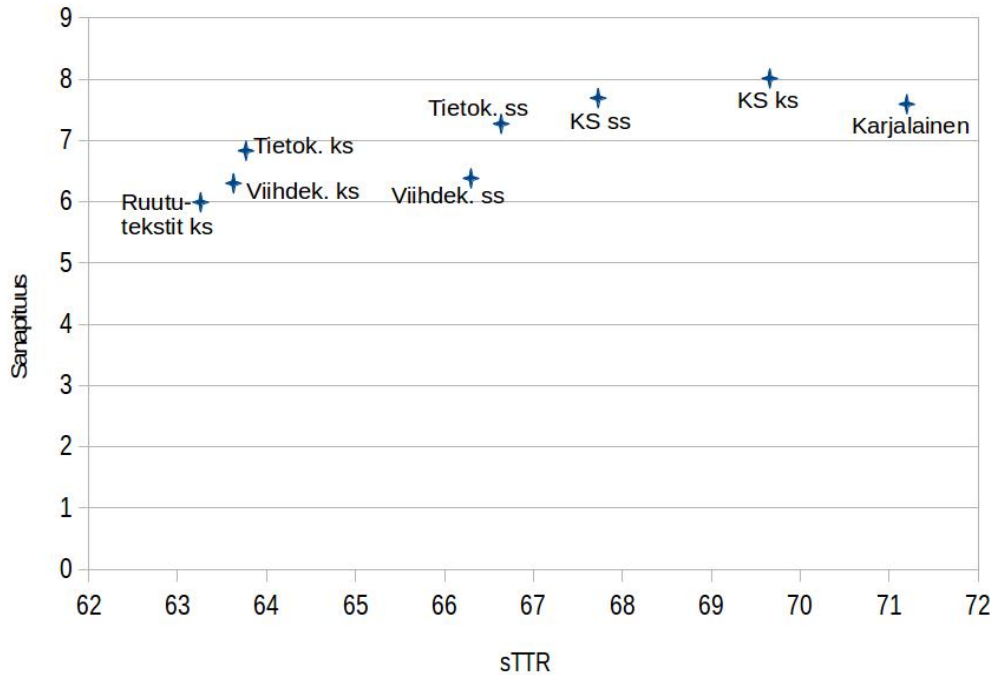
Voidaan olettaa, että enemmistösuomea ja vähemmistösuomea edustavat sanomalehtitekstit kielentävät eri sisältöjä johtuen kulttuurisista ja kielenkäyttötilanteiden eroista. Vähemmistösuomen käännettyjen ja ei-käännettyjen tekstien mahdolliset sisällölliset erot taas ovat vaikeammin ennakoitavissa. Molemmat tekstikorpukset on kuitenkin koottu yhden ja saman verrattain pienen kieliyhteisön tekstiaineistoista. Siksi odottaisi molempien tekstiaineistojen heijastavan kyseisen yhteisön elämän kannalta relevantteja sisältöpiirteitä riippumatta tekstin tuottamisen tavasta.

6 Analyysi ja tulokset

Sanapituutta pidetään yhtenä tekstin abstraktiuden ja vaikealukuisuuden indikaattorina (Mikk 2008), kieltenvälisestäkin, vaikka toki vain vertailtaessa tekstejä yhden kielen sisällä. Havaintojemme perusteella suomen kielen tekstilajien sanapituus asettuu yleensä niin, että tekstilajista riippumatta teksteissä on eniten (tilastollisesti moodi) 5-kirjaimisia sanoja; keskiarvona pituus voi kuitenkin vaihdella 5:n ja 8:n välillä. Tässä *Karjala Sanomien* sanapituutta ja sTTR:a siis verrataan *Karjalaiseen*, supi- ja käännös-suomen viihde- ja tietokirjallisuuteen sekä ruututeksteihin.

Vertailuaineistoihin nähden sanomalehtiteksti on huomattavan pitkäsanaisista (kaavio 1). Pitkäsanaisimpia ovat *Karjalan Sanomien* venäjämästä käännetty teksti (keskiarvo 8,01 merkkiä) ja suomeksi kirjoitettu teksti (7,69). *Karjalaisen* teksti on kolmanneksi pitkäsanaisinta (7,59). Sanomalehtiteksti sisältää paljon instituutioiden nimiä yms., joissa on paljon yhdyssanoja, mutta huomionarvoisinta tässä on se, että *Karjalan Sanomien* osalta pitkäsanaisin on nimenomaan käännetty tekstimuoto, kun se suomalaisilla tekstilajeilla on nimenomaan supisuomen tekstimuoto (viihdekirjallisuus: ss 6,38, ks 6,3; tietokirjallisuus: ss 7,27, ks 6,83). Ymmärrettävästi ruututekstien sanat ovat lyhimpiä (5,99). Teoreettisesti tärkeää on huomata, että eri tekstilajit ryhmittyvät lähekkäin tekstilajin mukaan eivätkä sen mukaan, ovatko ne supi- vai käännettyä suomea. Sanapituus on ensisijaisesti tekstilajiominaisuus: sanomalehtiteksti, viihdekirjallisuus ja tietokirjallisuus ryhmittyvät pareittain supi- ja käännös-suomeksi.

Kaavio 1: Osakorpusten sanapituudet ja sTTR (standardoitu type/token-suhde)



Mitä korkeammalla piste on kaaviossa, sitä abstraktimpaa ja vaikealukuisempaa (pitkäsanaaisempaa) teksti on, ja mitä enemmän vasemmalla piste on, sitä toisteisempaa teksti sanastoltaan on.

TTR sen sijaan lasketaan siis suhdelukuna, jossa korpuksen sananmuotojen absoluuttinen määrä jaetaan saneiden kokonaismäärällä. Tällaisella tunnusluvulla ei voi verrata kovin erimittaisia tekstejä keskenään, koska korpuksen kokonaissanamäärä on kaavassa jakajana. Standardoidussa versiossa (sTTR) tekstit jaetaan 1000 sanan segmenteiksi, joille jokaiselle lasketaan erillinen suhdeluku (jakaja on kaikille sama) ja koko korpuksen suhdeluku muodostuu niiden keskiarvosta. sTTR kuvaa siis suoraan tekstin sananmuotojen toisteisuutta: luku 60 tarkoittaa, että tekstin sananmuodoista 40 % (=100–60) on sellaisia, että ne esiintyvät useammin kuin kerran.

Aineistossa toisteisinta ovat luonnollisesti ruututekstit (avainsanaindeksi 63,26). Toisin kuin sanapituuden suhteen tekstin sanaston toisteisuus on ensisijaisesti käännöskielen ominaisuus, sillä eri tekstilajit ryhmittyvät sen mukaan, onko teksti supivai käännösuomea: ruututekstin jälkeen toisteisimpia ovat käännetyn suomen muodot (viihdekirjallisuus 63,63, tietokirjallisuus 63,77), kun taas selvästi sanastoltaan vaihtelevampia ovat supisuomalainen viihde- (66,30) ja tietokirjallisuus (66,64). Sanomalehtiteksti on rikkainta (Karjalainen 71,20), mutta jälleen huomionarvoista on se, että Karjalan Sanomien käännetty kielimuoto on rikkaampaa (69,66) kuin suomeksi kirjoitettu (67,73).

Karjalan Sanomien sanaston kvantitatiivisista piirteistä voidaan siis todeta, että sen kieli on sanastoltaan toisteisempaa kuin suomalainen sanomalehtiteksti, mutta toisaalta se on pitkäsanaaisempaa; erityisesti merkittävää on se, että nimenomaan *Karjalan Sanomien* supisuomi on toisteisempaa kuin käännösuomi, kun suhde yleensä on

päinvastainen. Muihin suomalaisiin tekstilajeihin verrattaessa voidaan todeta, että *Karjalan Sanomien* supisuomi on selvästi lähempänä suomalaista tietokirjallisuuden tekstiä kuin sanomalehtitekstiä; käännössuomi taas on nimenomaan lähempänä suomalaista sanomalehtitekstiä. – Se, että pitkäsanaisuus tässä suoraan liittyisi tekstin abstraktiuteen ja vaikealukuisuuteen, on mielestämme asetettava kyseenalaiseksi, ja olisi erikseen laadullisesti selvitettävä se, mihin *Karjalan Sanomien* pitkät sanat liittyvät. Tulos on kuitenkin mielenkiintoinen.

Sanaston frekvenssilistat nostavat hyvin yksinkertaisesti esiin tekstien yleisimmät sananmuodot. Seuraavassa vertaillaan *Karjalan Sanomien* suomeksi kirjoitettua kielimuotoa *Karjalaiseen* (ks. taulukko 2).

Taulukko 2: *Karjalan Sanomien* supisuomen ja *Karjalaisen* 25 yleisintä sananmuotoa (# tarkoittaa kaikkia numeroja)

| Karjalainen | KS-supisuomi |
|-------------|--------------|
| JA | JA |
| ON | ON |
| # | # |
| ETTÄ | KARJALAN |
| EI | ETTÄ |
| OLI | EI |
| MYÖS | OLI |
| OVAT | OVAT |
| SE | MYÖS |
| MUTTA | HÄN |
| OLE | SUOMEN |
| HÄN | SE |
| KUN | VUONNA |
| MUKAAN | OLE |
| KUIN | MUKAAN |
| SEN | VENÄJÄN |
| JO | PETROSKOIN |
| JOKA | JOKA |
| TAI | NYT |
| NIIN | SEN |
| SEKÄ | KERTOON |
| N | SANOO |
| NYT | MUTTA |
| VAIN | HE |
| OLLUT | PALJON |

25 yleisimmän sananmuodon sanalistat tuovat ensivaikutelmana esiin sen, että ne koostuvat paljolti samoista sanoista. *Karjalaiseen* verrattuna *Karjalan Sanomien* listalla esiintyy Karjalan tasavaltaan liittyviä paikannimiä ja verbimuotoja. Tarkempaa analyysia varten tarkastelimme 100, 200 ja 300 yleisimmän sanan sanalistoja niin, että luokittelimme sananmuodot nomineihin, verbeihin (sisältösanoja) ja funktiosanoihin.

Taulukot 3 ja 4 kokoavat sanaluokkien määrälliset jakaumat *Karjalan Sanomissa* ja *Karjalaisessa*.

Taulukko 3: Sisältö- ja funktiosanojen suhteellinen jakauma *Karjalan Sanomien* supisuomessa

| | <100 | <200 | <300 |
|--------------|--------|--------|--------|
| Verbit | 16,0 % | 15,0 % | 17,3 % |
| Nominit | 36,0 % | 48,5 % | 49,0 % |
| Funktiosanat | 48,0 % | 36,5 % | 33,7 % |
| | 100 % | 100 % | 100 % |

Taulukko 4: Sisältö- ja funktiosanojen suhteellinen jakauma *Karjalaisessa*

| | <100 | <200 | <300 |
|--------------|--------|--------|--------|
| Verbit | 17,0 % | 16,5 % | 18,7 % |
| Nominit | 21,0 % | 28,5 % | 32,7 % |
| Funktiosanat | 62,0 % | 55,0 % | 48,7 % |
| | 100 % | 100 % | 100 % |

Korpusten sananmuotojen frekvenssilistassa usein esiintyvät funktiosanat, esim. konjunktiot ja tietyt ajan adverbit sekä postpositiot, painottuvat listan alkuun, ja sen vuoksi niiden osuus vähenee, kun sanalista pitenee 100 yleisimmästä sananmuodosta 300:aan: *Karjalan Sanomissa* 48 %:sta 33,7 %:iin ja *Karjalaisessa* 62 %:sta 48,7 %:iin. Samoin kummassakin lehdessä nominien osuus vastaavasti kasvaa listan pidetessä, mutta verbien osuus pysyy lähes samana. Kun verrataan lehtiä keskenään, niin sanaluokkien jakaumat eroavat jo erittäin merkitsevästi toisistaan 200 sanan (χ^2 , $p=0,0001$) ja 300 sanan (χ^2 , $p=0,0001$) listoissa, ja lähes merkitsevästi myös 100 sanan (χ^2 , $p = 0,056$) listoissa: ero syntyy nominien suuresta osuudesta *Karjalan Sanomissa* ja funktiosanojen suuresta osuudesta *Karjalaisessa*.

Seuraavaksi tarkastellaan avainsanalistoja karjalansuomen ja supisuomen välillä. Taulukon 5 kolmeen osataulukkoon on koottu sanalistat niistä nominien, verbien ja funktiosanojen sananmuodoista, joita on tilastollisesti merkitsevästi (log-likelihood, $p<0,05$) enemmän toisessa kuin toisessa: jokainen taulu on jaettu kahtia niin, että vasemmalla ovat sananmuodot, joita on enemmän *Karjalan Sanomissa* kuin *Karjalaisessa*, ja oikealla päinvastoin.

Taulukko 5: Verbien, nominien ja funktiosanojen avainsanat: vasemmassa sarakkeessa *Karjalan Sanomien* ei-käännettyssä kielimuodossa ja oikeassa sarakkeessa *Karjalaisessa* ylliedustuvat sanat

| VERBIT | | NOMINIT | | FUNKTIOSANAT | |
|---------------|--------------|--------------|------------------|--------------|--------------|
| KS-ss > Karj | Karj > KS-ss | KS-ss > Karj | Karj > KS-ss | KS-ss > Karj | Karj > KS-ss |
| kertoo | olla | karjalan | joensuu | ja | n. |
| sanoo | tulisi | petroskoin | eu:n | jotka | mutta |
| olen | olisi | kielen | joensuun | nykyään | kun |
| opiskella | olevan | tasavallan | pohjois-karjalan | muttei | vaikka |
| auttaa | onkin | karjalassa | oy | yli | että |
| osallistui | arvioi | petroskoissa | kunnan | kovasti | jo |
| järjestettiin | ollut | venäjän | presidentti | joskus | ettei |
| oppivat | ollaan | ruplaa | matti | täten | niin |
| opiskelevat | saattaa | jelena | pekka | hyvin | ainakin |
| asuu | pitäisi | ks | tiistaina | harvoin | varsin |
| selittää | totesi | marina | lieksan | joissa | kyllä |
| auttavat | sanoi | kieltä | kuntien | nyt | vaan |
| järjestetään | nousi | suomea | maakunnan | siellä | vielä |
| omistettu | saisi | karjala | pääministeri | | kuin |
| myönnetään | jäi | tasavallassa | euroopan | | joten |
| pidettiin | arvelee | aleksandr | torstaina | | peräti |
| opiskelee | voida | suomen | kari | | ainakaan |
| pidän | pyritään | kižin | koko | | tosin |
| tutustua | tulee | lapset | jari | | selvästi |
| kehittää | lienee | petroskoihin | keskiviikkona | | hieman |

Kun avainsana-listoista otetaan huomioon kaikki sananmuodot (taulukossa 5 vain 20 merkitsevintä) ja kun nominien sanalistaista poistetaan erisnimet (*Petroskoi, Jelena, Aleksandr; Matti, Pekka*) ja instituutioiden nimet (*tasavalta; EU, kunta, pääministeri*) niin analyysi tuo esiin sanastoista sisällöllisesti sen, että

- nomineista Karjalan Sanomissa korostuvat *kielet, lapset, museot, koulut ja teatterit* ja Karjalaisessa viikonpäivät, *arkisto, poliisi, prosentit ja miljoonat*
- verbeistä Karjalan Sanomissa korostuvat *kertoa, sanoa, asua, opiskella, järjestää, auttaa, osallistua ja myöntää* ja Karjalaisessa *olla, tulla, arvioida, arvella, pyrkiä, todeta* ja erilaiset modaaliverbit (*saattaa, pitää, voida, saada*)
- funktiosanoista Karjalan Sanomissa korostuvat *ja*-konjunktio ja *täten*-adverbi, relatiivipronomini *joka* sekä tietyt ajanilmaukset (*nykyään, nyt, harvoin, joskus*) ja Karjalaisessa konjunktiot *mutta, kun, vaikka, että, ettei, vaan, joten, jos, tai, siten* sekä ajanilmaukset *jo, vielä, jälleen, yleensä* sekä adverbit *ainakin, ainakaan, varsin, tosin, peräti, ihan*.

Avainsana-analyysi *Karjalan Sanomien* supisuomen sanastosta paljastaa yhden mahdollisen syntaktisen piirteen funktiosanojen avulla eli sen, että sivulauseiden vähyys on sen erityinen piirre *Karjalaisen* supisuomeen verrattuna. Piirre varmistuu tietyksi vasta

laskemalla sivulauseiden suhteelliset osuudet, mutta erityisesti alistuskonjunktiot ovat aliedustuneita karjalansuomessa; rinnastuskonjunktiot *ja* ja *täten* taas ovat siinä yliedustuneita. Aiemmat tutkimukset ovat osoittaneet, että venäjänkielisissä lehtiteksteissä samoin kuin niistä tehdyissä käännöksissä virkkeet ovat huomattavasti pidempiä kuin suomenkielisissä rinnakkaisteksteissä (Vehmas-Lehto 2005: 66–67). Yhtenä syynä konjunktioiden runsaaseen käyttöön saattaisikin olla venäjän kielen antama malli, jonka mukaan lauseita samaan virkkeeseen rinnastuskonjunktion avulla. Suomessa aloitetaan mieluummin uusi virke.

Kun sama avainsana-vertailu tehdään *Karjalan Sanomien* käännösuomen ja *Karjalaisen* välillä, syntyy sanaston sisällöllisistä eroista hyvin samantyyppinen kuva (ks. taulukko 6) kuin yllä taulukosta 5. Se, että käännösuomen vertailuun kertyy määrällisesti vähemmän avainsanoja, johtuu *Karjalan Sanomien* käännöskorpuksen pienuudesta supisuomen korpukseen verrattuna.

Taulukko 6: Verbien, nominien ja funktiosanojen avainsanat: vasemmassa sarakkeessa *Karjalan Sanomien* käännösuomessa ja oikeassa sarakkeessa *Karjalaisessa* yliedustuvat sanat

| VERBIT | | NOMINIT | | FUNKTIOSANAT | |
|---------------|--------------|--------------|--------------|--------------|--------------|
| KS-ks > Karj | Karj > KS-ks | KS-ks > Karj | Karj > KS-ks | KS-ks > Karj | Karj > KS-ks |
| pidettiin | ei | karjalan | joensuun | ja n. | |
| järjestettiin | olisi | petroskoin | sillä | nykyään | kun |
| osallistuivat | olla | tasavallan | koko | jotka | mutta |
| osallistui | ollut | ks se | | yli | myös |
| omistettu | olevan | aleksandr | tämä | täten | jos |
| valmistaa | ole | petroskoissa | tässä | vaikka | |
| avattiin | onkin | piirin | oy | sillä | |
| osallistuu | | valeri | sitä | että | |
| kasvanut | | ruplaa | mitä | vielä | |
| rakennetaan | | louhen | itse | jo | |
| esiintyvät | | venäjän | sijaan | niin | |
| laulavat | | karjalassa | | kuin | |
| auttaa | | irina | | kyllä | |
| esitettiin | | sidorkin | | vaan | |
| ansioitunut | | hudilainen | | kuitenkin | |
| asuin | | kontupohjan | | kuitenkaan | |
| järjestetään | | sergei | | ettei | |
| paljastettiin | | sortavalan | | aivan | |
| harjoittavat | | tatjana | | vaan | |
| rakennettiin | | aleksei | | erityisesti | |

Laadullisesti käännössuomesta esiin nousevat avainsanat ovat samoja lemmoja (lekseemejä) kuin supisuomenkin vertailussa sillä erotuksella

- että listaus niistä verbimuodosta, joita on enemmän *Karjalaisessa* kuin *Karjalan Sanomissa*, sisältää vain olla-verbin muotoja
- että nominien listalle nousee oikeastaan pelkästään pronominiin *se, tämä, mitä* ja *koko* muotoja sekä adverbimainen (*sen*) *sijaan* ja lyhenne *oy*, kun paikannimi *Joensuu* jätetään taas tarkastelun ulkopuolelle.

Avainsanavertailussa kannattaa kohdistaa huomio myös syntaktisiin muotoryhmiin, koska sanamuodot kertovat myös morfosyntaktisista eroista, ja näin on etenkin verbien osalta. Kun tarkastellaan eroja *Karjalan Sanomien* supisuomen ja *Karjalaisen* välillä (taulukko 5), niin verbien muotoryhmistä korostuvat seuraavat piirteet:

- Karjalan Sanomissa yksikön ja monikon indikatiivin 3. persoonan muodot (*kertoo, sanoo, auttaa, opiskelevat, osallistui, oppivat*) ja passiivin preesens (*järjestettiin, järjestetään, myönnetään, pidettiin*)
- Karjalaisessa konditionaali (*tulisi, olisi, pitäisi, saisi*) ja imperfektin/preteritin 3. persoona (*arvioi, totesi, sanoi, nousi, jäi*) sekä vähäisessä määrin partisiipit (*olevan(sa), ollut, oltava*) ja infinitiivit (*olla, voida*).

Verbien vertailu paljastaa sen, että *Karjalan Sanomissa* kertovien väitelauseiden määrä on suhteellisen runsas eli yliedustunut, kun taas infinitiivirakenteiden (esim. modaaliverbit) ja lauseenvastikkeiden käyttöön liittyvät verbimuodot aliedustuneita. Näin muotoryhmiä tarkastelemalla paljastuu myös se, että *Karjalan Sanomien* supisuomi on siltä osin lähempänä lehden käännettyä suomea kuin *Karjalaisen* supisuomea.

7 Lopuksi

On oletettavaa, että nimenomaan venäläisen kulttuuriympäristön vaikutus on se, joka *Karjalan Sanomien* kielimuotoon avautuu tällaisen monipuolisen korpusanalyysin avulla, kun sitä verrataan *Karjalaisen* supisuomeen; samalla on tietysti todettava, että tällä se ei ole vielä todistettu. Venäjän kielen vaikutus Karjalan Sanomiin näkyy ensisijaisesti nominien yliedustumisena ja funktiosanojen aliedustumisena. Kun se liitetään siihen tulokseen että myös modaaliverbit aliedustuvat, samoin kuin tietyt merkitykseltään referoivat verbit, niin voidaan muodostaa hypoteesi, että suomalainen sanomalehtiteksti on ainakin argumentoivampaa kuin venäläinen lehtiteksti. Edelleen karjalansuomessa aliedustuvat sivulauseet ja erilaiset infinitiivirakenteet.

Korpusvertailu osoittaa, että karjalansuomen molemmat kielimuodot muistuttavat enemmän toisiaan kuin supisuomalaisista kielimuotoa, erityisesti sanastoltaan ja morfosyntaktisilta keinoiltaan. Jatkohypoteesien kannalta kiinnostavimpia olivat havainnot käännetyyn kielen sanastollisesta rikkaudesta verrattuna ei-käännettyyn kieleen,

eli se että tässä supisuomeksi kirjoitettu teksti on toisteisempaa kuin käännetty teksti. Nämä analyysitulokset luovat kiinnostavan teoreettisen lisän väitteisiin käännöskielen sanastollisesta yksinkertaisuudesta: vähemmistökielessä suhde on kenties päinvastainen. Kääntämisen rooli osana kielikontaktien ja vähemmistökielten problematiikkaa avaa uusia näkymiä tieteidenväliseen keskusteluun, johon voivat osallistua niin lingvistit kuin käännöstutkijatkin. Kauan jatkunut kääntämisen ”unohtaminen” tässä yhteydessä tuntuu vähintäänkin yhtä yllättävältä kuin aiemmin mainittu käännösten jättäminen kieliaineistojen kokoamishankkeiden ulkopuolelle. Käännöstutkijoiden ”kolmannella sukupolvella” on tilaisuus yhdistää kahden viime vuosikymmenen erillään kehittyneen tutkimusalueen saavutukset, toisin sanoen koota kielitieteestä kääntämiseen läheisesti liittyvää tietoa. Kiinnostavimpia kohteita ovat kielikontaktit, kaksikielisyys, vähemmistökielet ja toisen tai vieraan kielen omaksuminen. Kääntämisen käsitteen tarkasteleminen em. konteksteissa ja kääntämisen roolin saattaminen myös lingvistien tietoon nostaa oletettavasti esiin uusia relevantteja tutkimuskysymyksiä.

Tutkimusaineisto

Karjalainen-korpus. 1990-luvun suomen sähköinen tutkimusaineisto. Koonnut Jussi Niemi. Saatavilla: Kielipankki-palvelin, CSC – Tieteellinen laskenta Oy. <https://www.csc.fi/kielipankki>.

Karjalan Sanomat -korpus. Koonnut Anna Saikonen, Jukka Mäkisalo & Hannu Kemppanen.

Käännössuomen korpus. Koonnut Anna Mauranen. Joensuun yliopisto, Savonlinna 1997.

Kirjallisuuslähteet

Baker, Mona 1993. Corpus linguistics and translation studies. Implications and applications. Teoksessa: Mona Baker, Gill Francis & ElenaTognini-Bonelli (ed.) *Text and Technology. In Honour of John Sinclair*. Philadelphia & Amsterdam: John Benjamins, 233–250.

Baker, Paul 2006. *Using Corpora in Discourse Analysis*. London & New York: Continuum.

Becher, Viktor 2010. Abandoning the notion of “translation-inherent” explicitation. Against a dogma of translation studies. *Across Languages and Cultures* 11:1, 1–28.

Bernardini, Silvia & Adriano Ferraresi 2011. Practice, description and theory come together: Normalization or interference in Italian technical translation? *Meta* 56:2, 226–246.

Branchadell, Albert 2011. Minority languages and translation. Teoksessa: Yves Gambier & Luc van Doorslaer (ed.) *Handbook of Translation Studies. Volume 2*. Amsterdam & Philadelphia: John Benjamins, 97–101.

Cronin, Michael 1995. Altered States: Translation and minority languages. *TTR*, 8:1, 85–103.

Culpeper, Jonathan 2012. Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeares’s *Romeo and Juliet*. Teoksessa: Douglas Biber & Randi Reppen (ed.) *Corpus Linguistics. Volume IV. Methods and Applications*. London: SAGE, 81–109.

De Sutter, Gert, Patrick Goethals, Torsten Leuschner & Sonia Vandepitte 2012. Towards methodologically more rigorous corpus-based translation studies. *Across Languages and Cultures* 13:2, 137–143.

Eskola, Sari 2002. *Syntetisoivat rakenteet käännössuomessa. Suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmillä*. Joensuun yliopiston humanistisia julkaisuja 30. Joensuu: Joensuun yliopisto.

House, Juliane 2008. Beyond intervention: Universals in translation. *trans-kom. Zeitschrift für Translationswissenschaft und Fachkommunikation* 1:1, 6–19.

Jantunen, Jarmo H. 2011. Avainsana-analyysi annotoidun oppijankieliaineiston tutkimisessa: Alustavia havaintoja. Teoksessa: Esa Lehtinen, Sirkku Aaltonen, Merja Koskela, Elina Nevasaari & Mariann Skog-Södersved (toim.) *AFinLA-e. Soveltavan kielitieteen tutkimuksia*, n:o 3, 48–61.

Kemppanen, Hannu 2004. Keywords and Ideology in Translated History Texts: A Corpus-Based Analysis. *Across Languages and Cultures* 5:1, 89–106.

Kemppanen, Hannu 2008. *Avainsanoja ja ideologiaa: käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi*. Joensuun yliopiston humanistisia julkaisuja 51. Joensuu: Joensuun yliopisto.

Kenny, Dorothy 2000. Lexical hide-and-seek: looking for creativity in a parallel corpus. Teoksessa: Maeve Olohan (ed.) *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*. Manchester, UK & Northampton MA: St. Jerome, 93–104.

Kuusi, Päivi 2011. *Miksi näkökulma muuttuu käännöksessä? Eksplisiittistämisen ja normaalistamisen selitysvoima ja seuraukset*. Acta Universitatis Tamperensis 1638. Tampere: Tampere University Press.

Laviosa, Sara 2002. Europe in the making in translational and-non translational English. Teoksessa: Keith Harvey (ed.) *CTIS Occasional Papers. Volume 2*. Manchester: Centre of Translation and Intercultural Studies, UMIST, 85–95.

Laviosa-Braithwaite, Sara. 1996. *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation. Volume 1*. Department of Language Engineering. A thesis submitted to the University of Manchester Institute of Science and technology for the degree of Doctor of Philosophy.

Mahlberg, Michaela 2013. *Corpus Stylistics and Dickens's Fiction*. New York & London: Routledge.

Mauranen, Anna 2000. Strange strings in translated language. A study on corpora. Teoksessa: Maeve Olohan (ed.) *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*. Manchester, UK & Northampton MA: St. Jerome, 119–141.

Mauranen, Anna & Pekka Kujamäki (ed.) 2004. *Translation Universals – Do They Exist?* Amsterdam and Philadelphia: John Benjamins.

Madiba, Mbulungeni 2004. Parallel corpora as tools for developing indigenous languages of South Africa, with special reference to Venda. *Language Matters: Studies in the Languages of Africa* 35:1, 133–147.

- Mikk, Jaan 2008: Sentence length for revealing the cognitive load reversal effect in text comprehension. *Educational Studies* 34:2, 119–127.
- Nevalainen, Sampo 2004. Colloquialisms in translated text. Double illusion? *Across Languages and Cultures* 5:1, 67–88.
- Philips, Martin 1989. *Lexical Structure of Text. Discourse Analysis Monographs 12*. Birmingham: University of Birmingham.
- Pym, Anthony & Grzegorz Chrupala 2005. The quantitative analysis of translation flows in the age of an international language. Teoksessa: Albert Branchadell & Lowell Margaret West (ed.) *Less Translated Languages*. Amsterdam & Philadelphia, 27–38.
- Riionheimo, Helka, Leena Kolehmainen, Minna Kumpulainen, Lea Meriläinen, Pirkko Muikku-Werner & Esa Penttilä (toim.) 2014. Kontaktilingvistiikan, toisen kielen omaksumisen tutkimuksen ja käännöstieteen risteyksessä. *Virittäjä* 118:3.
- Scott, Mike 2000. Focusing on the text and its key words. Teoksessa: Lou Burnard & Tony McEnery (ed.) *Rethinking Language Pedagogy from Corpus Perspective. Volume 2*. Frankfurt: Peter Lang, 103–122.
- Scott, Mike 2008. *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software. Saatavissa: <http://www.lexically.net/wordsmith/index.html> [viitattu: 31.8.2015].
- Scott, Mike & Christopher Tribble 2006. *Textual Patterns. Key words and Corpus Analysis in Language Education*. Amsterdam & Philadelphia: John Benjamins.
- Sundelin, Apu 2000. *Taistelija, rakentaja, valistaja*. Petroskoi: Periodika.
- Tirkkonen-Condit, Sonja 2002. Translationese: a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target* 14:2, 207–220.
- Toury, Gideon 1985. Aspects of translating into minority languages from the point of view of Translation Studies. *Multilingua – Journal of Cross-Cultural and Interlanguage Communication* 4:1, 3–10.
- Vehmas-Lehto, Inkeri 2005. Rakennearsoja ja piilovirheitä. Kohti selkeämpää käsitystä käännösvirheistä. Teoksessa: Hilikka Yli-Jokipii (toim.) *Kielen matkassa multimediaan. Näkökulmia kääntämisen tutkimiseen ja opiskelemiseen*. Helsinki: Helsingin yliopisto, 49–74.
- Venuti, Lawrence 1998. Introduction. *The Translator* 4:2, 135–144.
- Zanettin, Federico 2012. *Translation-driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies*. Manchester, UK & Kinderhook (NY), USA: St. Jerome.

Kirjoittajien esittely ja yhteystiedot

Jukka Mäkisalo toimii kieliaineiden metodologian yliopistonlehtorina Itä-Suomen yliopistossa, ja hänen tutkimusaiheitaan ovat korpuslingvistiikka, audiovisuaalinen kääntäminen sekä kääntämisen kognitiiviset prosessit.

Sähköposti: jukka.makisalo (at) uef.fi

Hannu Kemppanen on venäjän kielen ja kääntämisen professori Itä-Suomen yliopistossa, ja hänen tutkimusaiheitaan ovat korpuslingvistinen käännöstutkimus sekä kääntäminen ja ideologiat.

Sähköposti: hannu.kemppanen (at) uef.fi

Anna Saikonen toimii suomen kielen opettajana Petroskoin valtiollisessa yliopistossa, ja hänen tutkimusaiheitaan ovat korpuslingvistiikka ja suomi vieraana kielenä -opetus.

Sähköposti: apetrova (at) petsu.ru