

Itämerensuomalaisten kielten keskinäistä ymmärrettävyyttä mittaavan kokeen suunnittelu, toteutus ja uutuusarvot

ALEKSI PALOKANGAS

Helsingin yliopisto

Tiivistelmä. Itämerensuomalaisten kielten keskinäisen ymmärrettävyyden tutkimus on osoittanut, että kielten äidinkieliset puhujat luottavat itämerensuomalaisten kielten samankaltaisuuteen pyrkiessään ymmärtämään sukukieltä. Tämä artikkeli käsittelee tutkimusta, joka tuo uusia näkökulmia aiheen tarkasteluun ja paneutuu eritoten aineiston keräämiseksi suunniteltuun kielikokeeseen. Se täydentää tutkimusaiheen aikaisempia menetelmiä, ja kokeen tavoitteet, suunnittelu, ominaisuudet ja kulku kuvataan tarkoin. Valotan koetekstin laadintaa ja koesanojen valintaa sekä esittelen järjestelmäni vastausten pisteyttämiseksi. Kuvaan pilottikokeiden kulun ja esittelen niiden tulokset. Tuloksista voidaan havaita, että koe soveltuu tutkimuskysymyksiini vastaamiseen. Koetulokset myös erosivat odotetusti, sillä eri taustamuuttujat, kuten kielitaito, ennakoivat erilaisia koetuloksia. Näitä havaintoja voidaan pitää merkkeinä siitä, että koe on saavuttanut sille asetetut tavoitteet.

Avainsanat: sukukielen ymmärtäminen; kielikoe; reseptiivinen monikielisyys; suomen kieli; viron kieli; vepsän kieli; liivin kieli

1. Tutkimuksen tausta ja tavoitteet

Itämerensuomalaisten kielten keskinäistä ymmärrettävyyttä on tutkittu jo melko paljon (ks. yhteenveto Härmävaara & Gooskens 2019). Erityisenä tutkimuskohteena on ollut viron ja suomen todellinen ja havaittu samankaltaisuus. Tutkimukset osoittavat kielten olevan niin samanlaisia, että niitä voi ymmärtää luettuna, eritoten muoto- ja lauseopillisia samankaltaisuuksia hyväksi käyttäen (Härmävaara & Gooskens 2019: 16). Tutkimuksissa on myös havaittu, että viron tai suomen äidinkielliset puhujat luottavat sukukielten samankaltaisuuteen pyrkiessään ymmärtämään suomea tai viroa (ks. esim. Kaivapalu 2005: 271; Kaivapalu & Muikku-Werner 2010: 83). Tutkimuksia on tehty yksittäisten sanojen, lauseiden ja kokonaisten tekstien kääntämisestä, mutta ei niitä tai usean itämerensuomalaisen kielen ymmärtämisen tutkimista yhdistellen.

Tutkimukseni käsittelee suomen ja viron keskinäistä ymmärrettävyyttä kirjoitetussa kielessä. Selvitän suomalaisille ja virolaisille koehenkilöille laatimani kokeen aineistosta kielellisiä tekijöitä, jotka vaikuttavat näiden kielten äidinkiellisten puhujien sukukielen ymmärtämiseen. Selvitän myös, miten yhden sukukielen taito hyödyttää muiden sukukielten ymmärtämistä, eli miten viroa tai suomea äidinkielenään puhuvat ja suomea tai viroa toisena kielenä osaavat ymmärtävät liiviä ja vepsää. Näin luon kokonaiskuvaan pohjoisten ja eteläisten itämerensuomalaisten kielten keskinäisestä ja sisäisestä ymmärrettävyydestä.

Uusia näkökulmia aikaisempaan tutkimukseen nähden ovat usean eri kielen ymmärtämisen samanaikainen tutkimus, eri taitotasolla lähdekieltä osaavien osallistuminen tutkimukseen ja heidän vastauksensa analysointi taitotasoin ja samojen koesanojen kääntäminen kolmessa eri yhteydessä: tutkin, miten sanojen ymmärtäminen vaihtelee, kun lähdekielen sana esitetään ensin yksittäin, sitten osana virkettä ja lopuksi tuossa virkkeessä osana lyhyttä tekstiä. Uusi avaus on myös vepsän ja liivin tutkiminen keskinäisen ymmärrettävyyden näkökulmasta. Vähemmistökieliä tutkimalla voidaan löytää jotain, mikä rikkoo

nykykäsityksiä keskinäisestä ymmärrettävyydestä, mitä on tutkittu eniten laajalevikkisten kielten osalta.

Aineiston keräämiseksi suunnittelin kielikokeen, jonka avulla selvitan, miten suomea tai viroa äidinkielenään puhuvat ymmärtävät sellaisia sukukielten sanoja, joista valtaosalla on etymologinen vastine heidän äidinkielessään. Kielikokeen avulla on tarkoitus vastata seuraaviin tutkimuskysymyksiin.

1. Miten valitsemieni etymologisten vastineiden havaitseminen ja ymmärtäminen muuttuu, kun vastineet esiintyvät yksittäin, lauseissa tai tekstien osana?
2. Miten suomalaisten ja virolaisten viron ja suomen kielitaitotaso vaikuttaa vastineiden havaitsemiseen ja ymmärtämiseen, kun ne esiintyvät yksittäin, lauseissa tai tekstien osana?
3. Miltä osin lähisukukielten ymmärtäminen on symmetristä ja miltä osin epäsymmetristä vastineiden esiintymisyhteys ja koehenkilöiden kielitaitotaso huomioiden?
4. Millaista on itämerensuomalaisten kielten pohjois- ja eteläryhmien sisäinen ja välinen keskinäinen ymmärrettävyys koehenkilöiden viron ja suomen kielitaitotaso huomioiden?

Kuvaan tässä artikkelissa aineiston keräämiseksi suunnittelemani ja alan aikaisempia menetelmiä täydentävää kielikoetta, sen laadintaa ja pilotointia. Pilotin avulla selvitin, että koe vastaa tarkoitustaan ja että voin sen tulosten avulla vastata tutkimuskysymyksiini. Artikkelissa valotan ensin aikaisempaa tutkimusta, mistä siirryn kielikokeen suunnittelun kuvailuun. Selvitän luvuissa 2–3, miten valitsin käännettävät koesanat ja laadin kokeessa käytettävät tekstit. Luvussa 3 esittelen järjestelmäni koevastausten pisteyttämiseksi, laatimani haastattelun ja kokeen rakenteen. Luvussa 4 kuvaan pilottikokeen ja käsittelen sen tuloksia. Luvuissa 4–5 kerron päätelmäni.

2. Katsaus aikaisempaan tutkimukseen

Itämerensuomalaisten kielten keskinäistä ymmärrettävyyttä ovat tutkineet kokonaisen tekstin tasolla ja suomalaisten viron ymmärtämisen osalta Kaivapalu ja Muikku-Werner (2010), ymmärtämisen epäsymmetrian osalta Kaivapalu ja Eslon (2011), ja yksittäissanojen tasolla ja suomalaisten viron ymmärtämisen osalta Muikku-Werner ja Heinonen (2012). Sekaannussanojen, lausetason ja suomalaisten viron ymmärtämisestä ovat kirjoittaneet Paajanen ja Muikku-Werner (2012) ja suomalaisten viron ymmärtämisen ja virolaisten suomen ymmärtämisen osalta Muikku-Werner (2014, 2015, 2016).

Tätä tutkimusta on koonnut yhteen Kaivapalu (2015), joka pureutuu artikkelissaan eritoten ymmärtämistestien tuloksissa ilmenneeseen keskinäisen ymmärrettävyyden epäsymmetriaan. Hän havaitsi, että suomalaiset vastaajat pärjäsivät virolaisia paremmin tekstinymmärrystä mittaavissa testeissä, kun taas virolaiset ymmärsivät paremmin irtosanoja. Yksittäisten sanojen ymmärtämistä mittaavissa testeissä virolaiset menestyivät suomalaisia paremmin etenkin semanttisesti ja ortografisesti samankaltaisten sanojen tunnistamisessa. Molemmat ryhmät käänivät heikosti niin kutsuttuja sekaannussanoja eli sanoja, joilla on kielissä samankaltainen kirjoitusasu mutta eri merkitys. Paajasen ja Muikku-Wernerin (2012) sekaannussanoja lausekonteksteissaan tarkastellut tutkimus kuitenkin osoittaa, että lausetason konteksti auttaa osallistujia päättelemään sanan merkityksen huomattavasti irtosanatestin sanoja useammin (ks. Härmävaara & Gooskens 2018: 18–19).

Syitä keskinäisen ymmärrettävyyden epäsymmetrialle on esitetty vasta varovasti. Sukukielelle altistumattomuuden on oletetusti havaittu olevan tärkein epäsymmetriaa selittävä syy. Esimerkiksi Härmävaaran ja Gooskensin (2019) irtosanojen ymmärtämistä mittaavassa tutkimuksessa vironkielisten osallistujien paremmat tulokset olivat pääasiassa selitettävissä virolaisten sukukielelle altistumisen suuremmalla määrällä (ks. myös Härmävaara 2022: 56–57). Lisäksi eri tutkimuksissa on nousut esiin muutama epäsymmetriaa selittävä kielellinen tekijä. Virossa

on ensinnäkin enemmän lainasanoja kuin suomessa, mikä mahdollistaa muihin kieliin tukeutumisen ymmärtämistehtäviä tehdessä. Äidinkielessä esiintyvä vaihtelu taas on noussut selittäväksi tekijäksi silloin, kun testattavan kielen sanalle löytyy lähimerkityksinen sana omasta äidinkielestä (ks. Kaivapalu 2015; Härmävaara & Gooskens 2019).

Härmävaara ja Gooskens (2019) havaitsivat irtosanojen keskinäistä ymmärrettävyyttä mittaavassa kokeessaan, että sukukielen kanssa hyvin vähän tekemisissä olleiden ryhmässä ymmärtämisen epäsymmetriaa ilmeni vain puhutussa muodossa esitettyjä testisanoja käännettäessä. Kielellisiksi syiksi epäsymmetriaan he esittivät muutamaa suomessa ja virossa toisistaan poikkeavaa piirrettä. Ne saattaisivat selittää Kaivapalun (2015) havaitseman eron siinä, miten virolaiset osallistujat tunnistavat ulkoasultaan ja merkitykseltään samankaltaisia sanoja paremmin kuin suomalaiset. Virolaiset osallistujat käänsivät paremmin sanoja, joista loppuvokaali on viron sanasta kulunut ja jossa se on suomen sanassa säilynyt (kuten *silm* – *silmä*). Viron sanojen ymmärtäminen vaikeutui myös silloin, kun sanassa oli suomesta puuttuva *õ*-ääne. Lisäksi suomalaisia osallistujia oli vienyt harhaan sanarakenne, jonka he tulkitsivat taivutusmuodoksi. Vastauksista esimerkiksi näkyi, että sana *kursus* 'kurssi' oli usein tulkittu inessiivimuotoiseksi sanaksi, kuten *korsussa*.

Selittääkseen ymmärtämisteisteissä ilmi tullutta epäsymmetriaa Kaivapalu ja Martin (2014, 2017) ovat pureutuneet samankaltaisuuden havaitsemisen kysymyksiin. Heidän kehittämänsä Index of Perceived Similarityn avulla on tarkoitus hahmottaa, miten kielellinen samankaltaisuus havaitaan. Kaivapalu ja Martin (2017) pyysivät koehenkilöitään ryhmittelemään pareittain esitettyjä suomen ja viron sananmuotoja niiden samankaltaisuuden mukaan kolmeen eri luokkaan. Koehenkilöiden tuli testissä ilmoittaa, ovatko parit heidän mielestään samanlaisia, jonkin verran samanlaisia vai keskenään erilaisia, ja he saattoivat myös perustella luokitteluitaan. Testisanat oli laadittu siten, että parittaiset sananmuodot kuuluvat muoto-opillisen vaihtelunsa puolesta neljään eri ryhmään: samanlaiset sanavartalot ja pääteainekset, samanlainen

sanavartalo ja erilainen pääteaines, erilainen sanavartalo ja samanlainen pääteaines, sekä erilaiset sanavartalot ja pääteaineokset.

Tutkimuksen havaintona oli, että suomalaiset näkivät virossa enemmän samankaltaisuutta suomen kanssa verrattuna virolaisten näkemyksiin viron ja suomen samankaltaisuudesta. Suomalaisten näkökulmasta viron sananmuotojen loppuista vaikutti ikään kuin puuttuvan jotain, kun taas virolaiset näkivät suomen sananmuotojen loppuissa tavallaan jotain ylimääräistä. Tämä asetelma toi suomalaisille etua, sillä tämä on tavalista lounaismurteissa ja muutoinkin suomen puhekielessä. (Kaivapalu & Martin 2017: 10–11, 19–21.)

Muikku-Werner (2014, 2015, 2016) on käsitellyt tutkimuksissaan fraseologisten yksiköiden ja kontekstin vaikutusta ymmärrettävyyteen. Vuonna 2015 julkaistussa artikkelissa Muikku-Werner esittelee viroa taitamattomien suomalaisten kykyä kääntää sellaisia fraseologisia yksiköitä, joiden jokin osa oli helpohkosti käännettävissä suomen avulla. Hän siis tutki asiayhteyden ja erilaisten vihjeiden osuutta sellaisten sanojen kääntämisessä, jotka eivät juuri muistuta suomen vastaavia sanoja. Havaintona oli, että erilaiset semanttiset sidokset edistävät tuntemattomien sanojen päättelyä, eli käännettävien sanojen lähellä olevat suomea muistuttavat sanat auttoivat päättelyssä (Muikku-Werner 2015: 210). Vastaavasti Muikku-Werner (2014: 115) havaitsi, että samankaltaisiakin sanoja voi olla vaikeaa ymmärtää käännettävänä olevan sanan lähellä olevien sanojen vaikutuksesta. Vuonna 2016 julkaistu artikkeli kuvaa sitä, miten virolaislukiolaiset käänsovät suomenkielisen tekstin ilman suomen opintoja. Tutkimus osoitti, että virolaiset tukeutuvat sanojen ulkoiseen samankaltaisuuteen pyrkiessään ymmärtämään suomea, mutta myös konteksti, ala- ja yläkäsitteet sekä syyt ja seuraussuhteet edesauttoivat ymmärtämistä (Muikku-Werner 2016: 311).

Keskinäistä ymmärrettävyyttä on tutkittu myös muiden itämerensuomalaisten kielten kuin suomen ja viron välillä. Alekseeva (2016) on tarkastellut vienankarjalan ja suomen keskinäistä ymmärrettävyyttä yksittäisten sanojen ymmärtämistä testaamalla. Hän tarkasteli yksittäissanojen ymmärtämistä ensin asiayhteydestään irrotettuina ja sitten

kokeensa toisessa osassa asiayhteydessään. Alekseeva (2016: 43) havaitsi, että sanojen kääntäminen asiayhteydessään paransi koehenkilöiden käännoستهتävien tuloksia selvästi (vrt. Paajanen & Muikku-Werner 2012).

Muikku-Werner ja Riionheimo (2018) ovat tutkineet tekstitason ymmärrettävyyttä livvinkarjan ja suomen välillä. Tekstinymmärtämistessissä suomen kielen yliopisto-opiskelijat käänsivät uutistekstin, ja lisäksi heitä pyydettiin kuvaamaan kirjallisesti käännostratkaisujaan. Kyseessä on sovellus ääneenajattelumenetelmästä, jossa osallistujat kertovat tehtävän herättämistä ajatuksista sekä kääntämisestään. Eniten ymmärtämistä edisti osallistujien mukaan sukulaissanojen samankaltaisuus. Myös kielten muoto- ja lauseopillisista yhteyksistä oli apua (Muikku-Werner & Riionheimo 2018: 285; vrt. Kaivapalu & Muikku-Werner 2010).

Muiden suomalais-ugrialaisten kielten keskinäistä ymmärrettävyyttä on tutkittu ilmeisen vähän. Eräs keskustelunavaus koskee sitä, millaista lähdekielen vaikutus on silloin kun lähdekieli (tai lähdekielet ovat) on muu(t) kuin suomi tai viro. Tämän ja muitakin aiheeseen liittyviä kysymyksiä ovat esittäneet sekä unkaria että suomea opiskelleet ja taustaltaan kaksikieliset Natalia Bolotova ja Nikolay Rakin, joista toisen äidinkieli on vepsä ja toisen komi – ja molempien toinen kieli venäjä. (Ks. Bolotova & Rakin 2007.) Vaikka suomalais-ugrialaisten kielten keskinäisen ymmärrettävyyden tutkimus on painottunut suomen ja viron tutkimiseen, asiaan on otettu kantaa muidenkin itämerensuomalaisten kielten suhteen eri teoksissa, artikkeleissa ja populääreissä yhteyksissä (ks. esim. Lehtinen 2008: 236; Mantila 2000: 11).

Kielten keskinäistä ymmärrettävyyttä on tutkittu myös muiden kieli-ryhmien kielten välillä. MICReLa-hankkeessa (Mutual Intelligibility of Closely Related Languages) on tarkasteltu 16:n indoeurooppalaisen kielen keskinäistä ymmärrettävyyttä, ja myös Härmävaaran ja Gooskensin (2019) tutkimus hyödynsi MICReLa-hankkeen menetelmiä. Oman testausmetodini taustainspiraationa on osin ollut MICReLa-hankkeessa myöhemmin sovellettu sanojen samankaltaisuuden ryhmittely (van Bezooijen & Gooskens 2007; ks. luku 3).

3. Kokeen suunnittelu

Aineiston keräämiseksi laadittu kirjallinen koe on tämän tutkimuksen keskeisin osa. Tässä luvussa käsittelen kokeen suunnittelun osa-alueita. Ensin kerron lähtökohdista, joiden pohjalta rakensin kokeen, ja kuvailen koetekstin. Sitten siirryn tekstiin upotettujen koeksanojen valintaan, tekstin laatimiseen, koevastausten pisteyttämiseen sekä kokeen rakenteen ja kulun esittelyyn. Testasin laatimaani koetta pilottitutkimuksella, jonka tuloksia esittelen. Lopuksi kerron tutkimushaastattelusta, johon osa vastaajista osallistui.

3.1. Kokeen lähtökohtia

Valitsin viron ja suomen lisäksi kokeeseen liivin ja vepsän, koska liivi kuuluu viron kera itämerensuomalaisten kielten eteläryhmään ja vepsä ja suomi taas pohjoisryhmään. Kokeen sanoista kolmanneksella onkin vastineensa vain etelä- ja pohjoisryhmässä. Tällä jaolla selvitän, millaista on etelä- ja pohjoisryhmien sisäinen ja välinen ymmärrettävyys ja hyödyttääkö sukukielen taito toisenkin sukukielen ymmärtämistä. Ymmärtämistä mitataan kokeessa käännoštehtävin, ja osallistujia pyydettiin myös perustelemaan käännoksiään.

Valitsin virolle pariiksi liivin ja suomelle vepsän, koska ne eroavat kielihistoriallisesti ja muutenkin selvästi virosta ja suomesta. Näin pystyn tutkimaan myös ryhmien sisäistä ymmärrettävyyttä, koska ryhmien sisäiset erot ovat kielihistoriallisesti suunnilleen samanlaiset (ks. mm. Kallio 2007: 243; Kallio 2024: 316–317). Pohjaan valintani siis kielimuotojen sisäisiin eroihin, polveutumiseen ja vuorovaikutukseen, joista ovat kirjoittaneet Kallion lisäksi muun muassa Sammallahti (1977: 131–132), Grünthal (2015: 53), O'Rourke & Pajusalu (2016: 69) ja Massinen (2022: 65–66). Kokeen keskiössä on lyhyt alun perin suomeksi kirjoittamani teksti muille tutkimuksen kielille käännettynä. Teksteistä on alleviivattu 17 käännettävää koeksanaa.

(1) Suomi

Isä on pitkä mies, mutta hän oli jo vanha. Huomasin sen, kun istuin tänään hänen rinnallaan ja hän taputti minua selkään. Isän otsa oli korkea, silmät siniset, ja hänellä oli pitkä parta ja pitkät hiukset. Hän oli kuin muinainen tietäjä – erityisesti kun hän soitti viisikielistä kanneltaan valkeassa paidassaan. Tuumin mitä antaa tällaiselle miehelle lahjaksi. Kello vai kultasormus? Ei hän tarvitse mitään. Ehkä odotan taas yhden vuoden.

Vepsä

Tat om pit'k mužik, no hän oli jo vanh. Minä homaičin sidä, konz ištuin tāmbej hänen rindal i hän täputi mindai sel'gha. Tatan oc oli korged, sil'mäd sinized, i hänel oli pit'k bard i pit'kad hibused. Hän oli kuti enzne tedai – sidä enamban konz hän soiti ičeze vižkeližel kantelel vauktas paidas. Dumain midä antta mugoižele mužikale lahjaks. Časud vai kuldsormuz? Hänele ei tarbiž nimidä. Aigoin varastan vól ühten voden.

Viro

Isa on pikk mees, aga ta oli juba vana. Märkasin seda, kui istusin täna tema kõrval ja ta patsutas mulle seljale. Isa otsmik oli kõrge, silmad sinised, ja tal oli pikk habe ja pikad juuksed. Ta oli kui muinasaegne teadja – eriti kui ta mängis oma viiekeelset kannelt, valge särk seljas. Mõtlesin, mida anda sellisele mehele kingiks. Kell või kuldsõrmus? Talle ei ole tarvis midagi. Ehk ootan jälle ühe aasta.

Liivi

Izä um pitkä miez, agä ta vól jõva vanä. Ma paņ sie tädöl, ku ma istiz tämpõ tām küoral ja ta patštiz min säлга. Izä võntsa vól kuordi, silmad – siņņizt, ja tämmõn vólđ pitkä abbõnd ja pitkäd ibükst. Ta vól nemē muinizäiga tiedaji – ižkiz siz, ku ta spēļiz eņš vidkēļizt kandlõ, vālđ serk sälgas. Ma mõtliz, mis ma ändaks škinkõks selļiz mien. Kiela agä kuldiz süormõks? Tämmõn äb uo vajäg migädid. Agä ma vóđlõb tegiž id äigast.

Lähdin testin laadinnassa liikkeelle sanastosta, jonka ymmärtämistä halusin testata. Luonnosteltuani ensimmäiset koeksanojen listat laadin koetekstin suomeksi ja sijoitin koeksanat tekstiin. Vironsin tekstin niin,

että jos sitä tehdessä olisi ilmennyt tilanteita, joissa olisi ollut kaksi yhtä luontevaa vaihtoehtoa käännökseksi, olisin käyttänyt sitä, mikä olisi ollut vastaavaa suomen ilmaisua muistuttava. Noudatin samaa linjaa myös kieliopin ja sanaston osalta. Kieliopin osalta tällaisia valintoja ei lopulta ollut tarve tehdä. Sovittelin käännöksiä toisiinsa varmistaakseni, että vastineiden merkitys ja käyttöala vastasivat toisiaan. Niinpä tekstiä muutettaessa yksittäisiä koesanojakin piti vaihtaa. Seuraavissa luvuissa kuvaan koesanojen valitsemisen periaatteita.

3.2. Koesanojen valinta

Jotta koe mittaisi keskinäistä ymmärrettävyyttä mahdollisimman laajasti, valitsin mukaan sanoja, joiden vaikeusaste ja muut ominaisuudet vaihtelevat paljon. Valinnassa ja ryhmittelyssä käytin muun muassa van Bezooijen ja Gooskensin (2007) kehittämää menetelmää. Se on neliportainen asteikko sanojen ryhmittelyyn sen mukaan, miten paljon lähdekielen sana eroaa kohdekielestä ja miten haastavaa sanan merkitystä on päätellä. Asteikko on laadittu ryhmittelemään sanoja, joilla on sama alkuperä ja suunnilleen sama merkitys. (van Bezooijen & Gooskens 2007: 255–256.)

Lisäksi noudatin sanojen valinnassa seuraavia periaatteita:

1. Tekstissä ei ole uusia lainasanoja, kuten viron ja suomen *auto* tai *maraton*, vaikka ne ovatkin näiden kielten jokapäiväistä sanastoa. Niiden yhtäläisyyksiä havaitessa koehenkilöt olisivat oletettavasti tukeutuneet muihinkin kuin itämerensuomalaisiin kieliin. Suomen ja viron keskinäistä ymmärrettävyyttä kartoittavissa tutkimuksissa on todettu, että viron ja suomen ymmärtämisen apuna käytetään myös muita kieliä, tosin myös sellaisissa tapauksissa, joissa sanalla on yhteinen lainattu alkuperä, kuten *röövida* 'ryövätä' (Kaivapalu 2015: 64–65). Uudet lainasanat eivät olisi siis välttämättä mitanneet vain itämerensuomalaisten kielten keskinäistä ymmärrettävyyttä. Muiden kielten apu ei painotu niin merkittävästi vanhojen lainasanojen tai jonkin

- yhteisen kantakielen sanoista periytyvien sanojen kääntämisessä, vaan niiden tulkitsemiseen käytettäneen apuna enimmäkseen suomea tai viroa (mts. 64–65).
2. Iso osa käännettävistä sanoista on kohtalaisen yleisiä, jotta tutkimuksessa olisi myös käytännönläheinen näkökulma. Yleiset sanat voivat antaa tietoa siitä, miten kielen ymmärtäminen sujuisi kokeen ulkopuolella, kuten matkustettaessa tai virolais-suomalaisessa työyhteisössä. Myös Härmävaaran ja Gooskensin (2019: 21–22, 51–54) tutkimuksessa tarkasteltiin yleisten sanojen kääntämistä. Siinä koeaineistona oli sata englannin yleisintä sanaa suomeksi ja viroksi käännettyinä. Niiden joukosta satunnaisotostettiin 50 sanaa kunkin koehenkilön kokeeseen (mts. 21). Satunnaisotos onkin yksi tapa välttyä siltä, että koe olisi painottunut sellaisiin heidän tutkimuksessaan olleisiin sanoihin, kuten *auto* tai *poliitika* ja *politiikka*.
 3. Käännettävistä 17 sanasta kymmenen on sellaisia, että niillä on etymologiset ja samaa merkitystä kantavat vastineet kaikissa tutkimuksen kielissä. Siksi esimerkiksi suomen *vain*, vepsän *vaiše*, viron *vaid* ja liivin *set* eivät soveltuneet kokeessa käännettäviksi sanoiksi, koska liivin sana poikkeaa etymologialtaan muista. (Ks. taulukot 1–2.)
 4. Vastaavasti seitsemällä sanoista ei ole merkittäviä etymologisia yhteyksiä pohjoisen ja eteläisen itämerensuomen välillä. Näilläkin sanoilla on etymologisia vastineita valittujen ryhmien välillä, kuten viron *habe* ja liivin *abbõnd*, jotka merkitsevät 'partaa'. Niiden vastineet suomen *ha(i)ven* ja vepsän *habin* 'nenäalus, viikset' ja *haben* 'partahaiven' merkitsevät kuitenkin hieman eri asioita. Jos sanoja, joilla ei ole samaa merkitseviä vastineita pohjoisen ja etelän välillä, olisi ollut vähemmän kuin seitsemän, olisi viron ja suomen taidon hyödyntämisen tarkkailu jäänyt vaillinaiseksi ja sanojen joukko ominaisuuksiltaan liian yksipuoliseksi.
 5. Käännettävien sanojen joukossa tuli olla tasaisesti sanoja van Bezooijen ja Gooskensin menetelmän asteikolta (ks. luku 3.3).

Esimerkiksi suomen ja viron *on* ja vepsän *om* ovat keskenään tasolla 0, mutta liivin *um* on tasolla 1, koska se eroaa selvästi muista ulkoasultaan. Toki näissä sanoissa on kaikissa yhtä monta kirjainta, ja ne alkavat takavokaaleja ja loppuvat nasaalikonsonantteja kuvaaviin kirjaimiin, jotka muistuttavat O:ta ja N:ää. Samoja kirjaimia niissä ei kuitenkaan ole.

6. Yksi aineiston keräämisen tavoite on tutkia keskinäistä ymmärrettävyyttä erilaisia sanoja käännettäessä. Siksi koesanat noudattavat tutkittavien kielten tavanomaista sanaluokkien suhdetta (ks. Heikkinen ym. 2001; ks. Tartu Ülikooli arvutilingvistika uurimisrühm 2019a, 2019b). Tämäkin liittyy käytännönläheisyyden liittämiseen tutkimukseen.
7. Halusin sanoissa olevan muoto- ja äänneopillista vaihtelua, kuten suomen, vepsän, viron ja liivin vastineissa: *yhden, ühten, ühe* ja *īd*. Niiden perusmuodot ovat *yksi, üks, üks* ja *ikš*, joten juuri taivutus tuo sanojen erot esiin. Toisena esimerkkinä mainittakoon sanat *paidassaan, paidas, särk* ja *serk*, joista ainoastaan suomen sana saa omistusliitteen ja eteläisen itämerensuomen sanat jäävät lauseopillisista syistä ilman taivutuspäätettä. Kuten esimerkit osoittavat, sanojen samannäköisyyteen, mikä on tärkeää kirjoitetun kielen ymmärtämistä tutkittaessa, vaikuttavat niin tavarakenteet kuin oikeinkirjoitussäännötkin.

3.3. Koesanojen luokittelu

Van Bezooijen ja Gooskensin (2007) kehittämän kielten samankaltaisuuden asteen kuvaamisen menetelmän ensimmäinen taso on 0 eli helpoin ymmärtämisen taso. Sillä lähdekielen ja kohdekielen suhde on täysin läpinäkyvä. Tällaisia sanoja ovat esimerkiksi suomen *ja* ja viron *ja*, joilla ei ole kirjoitusasuissa¹ mitään eroa ja joiden merkitys on sama.

¹ Kirjoitan tässä yhteydessä kirjaimista, en äänneistä, vaikka kirjaimet kuvaavatkin äänneitä, koska tutkimuksessa on kyse keskinäisestä ymmärrettävyydestä kirjoitetussa kielessä. On kuitenkin havaittu, että koehenkilöt käyttävät tällaisissa kokeissa

Tällä tasolla eroa voi kuitenkin olla jonkin verran, kuten suomen ja viron sanoissa *isä* ja *isa*. Sanojen kirjoitusasujen ero on vähäinen, ja ne lausutaan keskenään hyvin samantapaisesti. (Ks. van Bezooijen & Gooskens 2007: 256.)

Asteikon seuraava taso on 1. Siinä lähdekielen ja kohdekielen suhde on melko läpinäkyvä. Yleensä ero sanojen välillä on tasolla 1 yhden kirjaimen verran, mutta se voi olla enemmänkin. Tällaisia sanoja ovat esimerkiksi suomen *hauki* ja viron *haug*. Tasolla 2 sanojen suhde on jo melko läpinäkymätön, eli sanat muistuttavat toisiaan vain vähän. Yleensä ero on jo useita kirjaimia, kuten suomen ja viron sanoissa *palaa* ja *põleda*. Viimeinen taso on 3, jossa kohdekielen sanat eroavat niin paljon lähdekielestä, että niiden ymmärtäminen lähdekielen avulla on käytännössä mahdotonta. Esimerkiksi suomen ja viron *herätä* ja *ärgata* ovat tällaisia. (Ks. van Bezooijen & Gooskens 2007: 256.)

Tutkimuksessani käytettävät samaa alkuperää olevat sanat on esitetty taulukossa 1, jossa van Bezooijen ja Gooskensin menetelmän määrittelemä sanojen samankaltaisuuden taso on kirjoitettu kunkin sanan viereen suluissa. Suomesta ja virosta on ilmoitettu samankaltaisuuden taso suhteessa toisiinsa, kun taas vepsästä ja liivistä on ilmoitettu samankaltaisuuden taso ensin suhteessa suomeen ja toisena suhteessa viroon.

Taulukossa 1 on ilmoitettu myös sanojen esiintymistaajuudet, jotka viittaavat sananmuotojen esiintymistiheyteen Kotimaisten kielten keskuksen Parole-korpuksessa, jossa on noin 17,6 miljoonaa sanetta (ks. Kirjoitetun suomen kielen sanojen taajuuksia), ja Tarton yliopiston Eesti keele koondkorpuksen tasapainotetussa osiossa, jossa on noin 15,0 miljoonaa sanetta (ks. Tartu Ülikooli arvutilingvistika uurimisrühm 2024). Valitsin koesanat valitsemalla ne taajuussanaston yleisemmästä päästä luvussa 3.2 määrittelemieni ominaisuuksien mukaan.

päättykeininä myös käännettävien sanojen mielessä tai ääneen lausumista (ks. Kaivapalu & Muikku-Werner 2010: 84; Palokangas 2018: 60–61).

TAULUKKO 1. *Samaa alkuperää olevien koesanojen ominaisuuksia*

Suomi	Vepsä	Viro	Liivi	Esiintymistaajuus		Sana- luokka
				suomessa	virossa	
<i>on</i> (0)	<i>om</i> (0, 0)	<i>on</i> (0)	<i>um</i> (2, 2)	2.	2.	verbi
<i>pitkä</i> (1)	<i>pit'k</i> (0, 1)	<i>pikk</i> (2)	<i>pitkā</i> (0, 1)	717.	733.	adjektiivi
<i>vanha</i> (0)	<i>vanh</i> (0, 0)	<i>vana</i> (1)	<i>vanā</i> (1, 0)	502.	252.	adjektiivi
<i>istuin</i> (1)	<i>ištuin</i> (0, 1)	<i>istusin</i> (2)	<i>istiz</i> (2, 2)	5 031.	2 849.	verbi
<i>tänään</i> (1)	<i>tämbei</i> (2, 3)	<i>täna</i> (1)	<i>tämpõ</i> (3, 3)	329.	260.	adverbi
<i>otsa</i> (1)	<i>oc</i> (3, 2)	<i>otsmik</i> (2)	<i>võntsa</i> (3, 3)	21 846.	–	substan- tiivi
<i>silmät</i> (0)	<i>sil'mäd</i> (0, 0)	<i>silmad</i> (1)	<i>silmad</i> (0, 0)	1 150.	396.	substan- tiivi
<i>hiukset</i> (2)	<i>hibused</i> (2, 3)	<i>juuksed</i> (3)	<i>ibükst</i> (3, 3)	4 105.	1 639.	substan- tiivi
<i>yhden</i> (1)	<i>ühten</i> (1, 2)	<i>ühe</i> (1)	<i>id</i> (3, 3)	221.	108.	numeraali
<i>jo</i> (2)	<i>jo</i> (0, 2)	<i>juba</i> (2)	<i>jõva</i> (3, 2)	19.	29.	adverbi

17 käännettävästä suomen sananmuodosta 6 on 750 yleisimmän suomen sanan joukossa (ks. Taulukko 1). Lemmattuina ne ovat vielä yleisempiä. Huomioin myös muiden kuin käännettävien sanojen esiintymistiheyden, jotta nekin olivat melko yleisiä.

Van Bezooijenin ja Gooskensin (2007) menetelmän lisäksi huomioin koetta laatiessani sanojen Levenšteinin etäisyyden, josta käytetään myös nimitystä muokkausetäisyys (Mustonen 2014). Taulukko 2 esittää nämä etäisyydet. Kielitieteessä ja muuallakin käytettävä muokkausetäisyys kertoo, kuinka monta lisäystä, poistoa, korvausta tai merkkien keskinäistä

vaihtoa on vähintään tehtävä, jotta kaksi merkkijonoa muuttuu samantyyppiseksi (Borin 2013: 10). Sanojen samankaltaisuutta määriteltäessä siinä on kuitenkin puutteensa. Se ei huomioi tarkkeita tai vokaaleiden tai konsonanttien läheisyyttä, kuten *u:n* ja *o:n* tai *m:n* ja *n:n* tapauksissa. Niinpä esimerkiksi Heeringa kollegooneen (2014: 109) laski tarkkeellisten kirjaimien (kuten ero kirjaimien *e* ja *é* tai *è* ja *é* välillä) eroksi 0,3 korvausta tavanomaisen yhden korvauksen sijaan. Tämänkin kokeen aineistossa on sanoja, jossa on tarkkeita, kuten suomen ja liivin vastineet *pitkä* ja *pitkã*.

TAULUKKO 2. *Samaa alkuperää olevien koesanojen muokkausetäisyydet. Suluiissa etäisyys niin, että tarkkeen korvaus- tai poistoarvo on 0,3. Myös Ä:n tarkkeet on huomioitu*

Koesanat				Muokkausetäisyys				
suomi	vepsä	viro	liivi	suomi-vepsä	suomi-viro	suomi-liivi	viro-vepsä	viro-liivi
<i>on</i>	<i>om</i>	<i>on</i>	<i>um</i>	1	0	2	1	2
<i>pitkä</i>	<i>pit'k</i>	<i>pikk</i>	<i>pitkã</i>	2 (1,3)	2	1 (0,3)	2 (1,3)	2
<i>vanha</i>	<i>vanh</i>	<i>vana</i>	<i>vanã</i>	1	1	2 (1,3)	1	1 (0,3)
<i>istuin</i>	<i>ištuin</i>	<i>istusin</i>	<i>istiz</i>	1 (0,3)	1	3 (2,3)	2 (1,3)	4 (3,3)
<i>tänään</i>	<i>tãmbei</i>	<i>tãna</i>	<i>tãmpõ</i>	4	3 (2,3)	4	4	3
<i>otsa</i>	<i>oc</i>	<i>otsmik</i>	<i>võntsa</i>	3	3	3 (2,3)	5	6 (5,3)
<i>silmät</i>	<i>sil'mäd</i>	<i>silmad</i>	<i>silmad</i>	2 (1,3)	2 (1,3)	3 (1,6)	2 (0,6)	1 (0,3)
<i>hiukset</i>	<i>hibused</i>	<i>juuksesed</i>	<i>ibükst</i>	3	3	4 (3,3)	4	5 (4,3)
<i>yhden</i>	<i>ühten</i>	<i>ühe</i>	<i>id</i>	2	3	4	2	3
<i>jo</i>	<i>jo</i>	<i>juba</i>	<i>jõva</i>	0	3	3 (2,3)	3	2

Yllä on esitetty muokkausepäilyksineen ne kokeen sanat, joilla on yhteinen alkuperä ja merkitys. Kuten epäilyksistä ilmenee, sanoissa on eroja monipuolisesti. Osa sanoista on kauttaaltaan varsin samanlaisia, kuten *on*-sana vastineineen. Osassa ero on pieni ryhmän sisäisesti, kuten *vanha*- ja *vana*-sanojen vastineet osoittavat. Vastaavasti osalla erot ovat merkittäviä, kuten *hiukset*-sanan ja sen vastineiden tapauksessa. Paikoin erot ovat pienimpiä juuri viron ja suomen välillä. Tästä esimerkkinä ovat viron ja suomen sanat *on* ja *on*.

3.4. Kokeen pisteytysjärjestelmä

Laadin koetta varten pisteytysjärjestelmän. Se perustuu osin maisterintutkimukseen (Palokangas 2018: 46–51) kehittämäni järjestelmään. Olen täydentänyt järjestelmää, jotta se mahdollisimman tarkasti kuvaisi ymmärtämistä asteittaisena ilmiönä, joka ei siis ole luonteeltaan kaksiarvioista. Järjestelmällä voi myös vertailla ymmärrettävyyttä edellä mainituista näkökulmista. Se siis täydentää van Bezoijen ja Gooskensin menetelmää, jonka asteikko on jokseenkin väljä. Muokkausepäilyteen nähden pisteytysjärjestelmä on huomattavasti joustavampi mutta silti tarkempi. Kussakin koeosiossa käännetään 17 sananmuotoa, eli yhden osion enimmäispistemäärä on 17,000 pistettä.

Pisteytysjärjestelmä perustuu ajatukseen, jossa tärkeintä on merkityksen ja muodon yhtäaikaista kääntäminen oikein. Sen lisäksi vastaajia palkitaan etymologisten yhteyksiin perustuvan samuuden havaitsemisesta, eritoten, jos se johtaa vastaajaa ymmärtämisen jäljille. On huomionarvoista, että etymologiset yhteydet eivät välttämättä ole vastaajille tärkeitä. Selvää lienee, että yhteyksien havaitseminen ei tietenkään aina tuota oikeaa käännöstä, mutta se voi johtaa siihen. Koska valtaosaa vastauksista ei ole perusteltu, ei aina ole mahdollista tietää, onko juuri etymologinen yhteys johtanut oikeaan tai väärään vastaukseen tai osittaiseen ymmärtämiseen.

Yhdestä käännöksestä annettava täysi pistemäärä on 1,000 pistettä, esimerkiksi vastaukseksi annettu ”parta” viron sanan *habe* käännökseenä.

1,000 pistettä saa myös oikean vastauksen synonyymista, esimerkiksi ”laup” suomen sanan *otsa* käännöksenä. Sen vastineita ja käännöksiä viroksi ovat myös ’otsaesine, otsmik’ ja ’ots’. Täydet pisteet saa myös yksittäisiä sanoja käännettäessä, kun sanan kirjoitusasu merkitsee useaa eri sanaluokkaa, kuten viron *kõrval* ’vierellä’, ’vieressä’ (adverbi), korvalla (adverbi ja substantiivi), ja sijamuotoa, kuten viron lekseemi *vana*, jonka taivutusmuodot nominatiivissa, akkusatiivissa, genetiivissä ja partitiivissa ovat *vana*, *vana*, *vana* ja *vana*. Täydet pisteet sai siis vastauksista ”korvalla” ja ”vanhaa”. Eihän koehenkilö voinut tietää, että sanat olivat tällä kertaa asiayhteyksistä, joissa ne merkitsivät ’vierellä’ ja nominatiivimuotoa ’vanha’. 1,000 pistettä sai, jos sanan merkitys oli käännetty oikein, mutta sen sijamuoto väärin, kuten ”vieressä” sanan *kõrval* käännöksenä.

0,875 pistettä annettiin lause- ja tekstiosioissa tietyistä oikeista vastauksista oikeassa taivutusmuodossa, josta kuitenkin puuttui omistusliite. Tällainen oli esimerkiksi vastaus ”vierellä, vieressä” sanan *kõrval* (ks. luku 3.1) käännöksenä. Oikea käännöshän olisi ”vierellään”, ”vieressään” tai ”rinnallaan”.

0,750 pistettä myönnettiin oikeasta vastauksesta väärässä taivutusmuodossa, esimerkiksi ”istun” liivin verbin *istiz* ’istuin’ käännöksenä. Tämä pistemäärä annettiin myös vastauksesta, joka oli oikeassa taivutusmuodossa mutta jonka merkityksessä oli pieni mutta ei erityisen merkittävä ero. Tällainen vastaus oli esimerkiksi yksittäisiä sanoja käännettäessä ”tämä” viron sanan *ta* (’hän’) käännöksenä, sillä sanoilla on etymologinen yhteys ja *ta*-sanaa voi käyttää myös merkityksessä ’se’ sen ensisijaisen merkityksen ’hän’ ohella. Suomessa *tämä* ja *se*-sanojen merkitysero on varsin pieni, ja virossa *ta*- ja *see*-sanat saavat myös merkityksiä, jotka vastaavat suomen sanan *tämä* merkitystä. Vastaus, jossa oli kirjoitusvirhe ja josta ei voinut olla täysin varma, mitä vastaaja sillä tarkoitti, sai 0,750 pistettä. Tällainen vastaus oli esimerkiksi ”yhde” viron *ühe* ’yhden’ käännöksenä.

0,625 pistettä sai, jos sanoja ei käännetty yksittäin ja vastauksen muoto oli oikea ja etymologia läheinen, mutta merkitys eriävä. Tällaisia olivat esimerkiksi virolaisen koehenkilön käännös, ”rinnal” ’rintakehällä,

rinnuksilla, vepsän *rindal*-sanasta. Myös rinnasteisista merkityksistä sai tämän verran pisteitä, kuten liivin *kūoral*-sanan, 'vierellä, vieressä, rinnalla', käännöksestä "kera" tai liivin sanan *jōva* 'jo' käännöksestä "jopa".

0,500 pistettä sai vastauksesta, joka oli lähellä oikeaa vastausta, mutta väärässä sanaluokassa, esimerkiksi substantiivi "istuin" liivin verbin *istiz* 'istuin' käännökseenä. Lause- ja tekstiosioissa oikeaan vastaukseen rinnastuvat paikan ilmaukset tuottivat 0,500 pistettä, kuten virolaisen vastaus "süles" 'sylissä' suomen adverbien *rinnallaan* käännökseenä. Käännettävän sanan yläkäsitteeksi käännettyt sanat samassa taivutusmuodossa tuottivat 0,500 pistettä, kuten "kosketti" liivin sanan *patštiz* 'taputti' käännökseenä. Käännös, jossa oli merkittävä etymologinen yhteys ja joka yhdistyi osittaiseen samaan merkitykseen käännettävän sanan kanssa, oli 0,500 pisteen arvoinen. Esimerkiksi viron "ajast" 'ajasta' ja suomen "aikaa" liivin sanan *āigast* 'vuoden' käännöksinä tuottivat 0,500 pistettä.

Jos merkitysoyhteyden keralla ei ollut etymologista yhteyttä, kuten liivin sanan *tämpō* 'tänään' vastauksessa "tuolloin", pisteitä sai 0,375. Niin oli myös, jos molemmat yhteydet olivat, mutta vastaus oli eri sanaluokassa, esimerkkinä adverbi "aikaisin" liivin substantiivin *āigast* 'vuoden' käännökseenä. Samat pisteet sai, jos vastaus oli oikean käännökseen yläkäsite, mutta väärässä taivutusmuodossa, kuten "koskettaa" liivin sanan *patštiz* 'taputti' käännökseenä. 0,375 pistettä sai, jos vastaus ja käännettävä sana olivat interjektioomaisia ja niillä oli etymologinen yhteys, mutta ei juuri muuta yhteyttä. Virolaisen vastaus "ju" '-hAn' sanan *jo* käännökseenä tuotti nämä pisteet.

Pisteitä sai 0,250, jos koehenkilö oli vastannut väärin, mutta vastaus vielä jotenkuten hipoi oikean vastauksen merkitystä. Tällainen vastaus olisi voinut olla esimerkiksi "sijainti" sanan *kōrval* 'rinnalla' käännökseenä, sillä "sijainti" on kohdesanan yläkäsite, mutta eri sanaluokassa ja taivutusmuodossa.

0,125 pistettä saavutti, jos vain vastauksen sijamuoto ja sanaluokka olivat oikein, kuten virolaisen vastauksessa "kuulsin" 'kuulin' vepsän verbin *ištuin* 'istuin' käännökseenä tai virolaisen vastaama "hobused" 'hevokset', vepsän *hibused* 'hiukset', käännökseenä. Myös kaukainen

etymologinen yhteys ja sen oheen annettu väärä taivutusmuoto tuottivat nämä pisteet tapauksessa, jossa sekä käännettävässä sanassa (*taputti*) että vastauksessa (”tapetud” ’tapetut’) oli kyse menneestä ajasta. Nämä pisteet sai myös, jos vastaus ja käännettävä sana olivat interjektiomaisia ja niillä oli kaukainen etymologinen yhteys, mutta ei juuri muuta yhteyttä, esimerkiksi virolaisen vastaajan ”ja” ’ja’ vepsän sanan *jo* ’jo’ käännöksenä. Jos koehenkilöllä oli jokin käsitys tai arvio käännettävän sanan merkityksestä, mutta hän käänsi sen väärään sanaluokkaan, annettiin pisteitä tämän verran. Tällainen käänнос oli esimerkiksi ”kaikki” liivin sanan *īd* ’yhden’ käännöksenä, sillä ne molemmat kuvaavat määrää.

3.5. Kokeen rakenne

Koe on suunniteltu toteutettavaksi siten, että suomalaiset koehenkilöt kääntävät ensin liiviä, sitten viroa ja lopuksi vepsää; virolaiset ensin vepsää, sitten suomea ja lopuksi liiviä. Koehenkilöt siis aloittavat kokeen kielestä, joka on kielihistoriallisesti kaukaisin heidän äidinkieleensä nähden. Koe loppuu kieleen, joka on kielihistoriallisesti lähinnä heidän äidinkieltään. Pyrin näin ehkäisemään sitä, että lähimmän sukukielen koeaineisto johdattelisi vastaajia kokeen edetessä kielestä toiseen, ja kokeen on tarkoitus edetä mahdollisimman pitkälle ennen kuin koehenkilöt havaitsevat kokeen tekstin olevan kaikilla kielillä sisällöltään sama. Pyrin tällä siihen, että koevastauksissa ilmenisi mahdollisimman vähän koeaineiston pohjalta oppimista.


Kokeen rakenne on esitetty taulukossa 3. Koehenkilöiden tuli kääntää alleviivatut sanat aluksi yksittäin, sitten osana virkettä ja lopuksi näissä virkkeissä osana tekstiä. Kolmasosa vastauksista oli perusteltava, jotta vastaajien päättelyä pystyy tarkastelemaan kokonaisvaltaisesti. Tämän jälkeen koe etenee seuraavaa kieltä koskeviin tehtäviin. Kysymykset esitetään samassa järjestyksessä kaikille vastaajille.

TAULUKKO 3. *Kokeen kulku*

Suomalaiset		Virolaiset	
Liivi, osat 1.1–1.3	Yksittäiset sanamuodot	Vepsä, osat 1.1–1.3	Yksittäiset sanamuodot
	Sanamuodot lauseissa		Sanamuodot lauseissa
	Sanamuodot tekstissä		Sanamuodot tekstissä
Viro, osat 2.1–2.3	Yksittäiset sanamuodot	Suomi, osat 2.1–2.3	Yksittäiset sanamuodot
	Sanamuodot lauseissa		Sanamuodot lauseissa
	Sanamuodot tekstissä		Sanamuodot tekstissä
Vepsä, osat 3.1–3.3	Yksittäiset sanamuodot	Liivi, osat 3.1–3.3	Yksittäiset sanamuodot
	Sanamuodot lauseissa		Sanamuodot lauseissa
	Sanamuodot tekstissä		Sanamuodot tekstissä

Testissä jokainen koehenkilö kääntää kullakin kielellä 17 sanaa. Tätä suurempi sanamäärä pitkittäisi koetta keskimäärin yli kahteen tuntiin ja siten voisi heikentää vastaamisen laatua. Tätä lyhyempi koe puolestaan supistaisi kerättävän aineiston kokoa. Kokeessa käytetyt sanamuodot ovat kaikille vastaajille samat. Kukin sanamuodoista esiintyy jokaisen koehenkilön testipatterissa kolmessa eri asiayhteydessä eli yksittäin, lauseessa ja tekstissä muodostaen 51 tehtävänkohtaa per kieli. Näin voin tutkia, miten sanan vieressä ja lähellä olevat sanat vaikuttavat ymmärtämisen prosessiin ja erilaisiin vastauksiin. Kyseistä ilmiötä kutsutaan pohjustamiseksi, missä yksi tai sitä useampi sana pohjustaa toisen osan eli kohdesanan ymmärtämistä, jos tuo pohjustinsana on tunnistettavissa esimerkiksi äidinkielen perusteella (Muikku-Werner & Jantunen 2020: 185). Tehtäviä on kolmesta eri kielestä, eli vastauksia tulee antaa yhteensä 153.

Jotta tiettyjen kokeen lopussa annettavien vastausten laatu ei kärsisi koehenkilöiden väsymyksen ja kognitiivisen rasituksen vuoksi, kysymykset on tarkoitettu esittämään kunkin osion sisällä satunnaistetussa järjestyksessä. Se ei kuitenkaan ollut mahdollista pilotissa käyttämälläni Qualtrics-ohjelmalla. Valtaosassa kysymyksiä oli 40 sekuntia vastausaikaa, kuten kuvasta 1 näkyy. Ajan kulku näkyi vastaajille. Tällaisia kysymyksiä oli 106, kun vastauksia tuli yhteensä antaa 151.



34

Käännä alleiviivattu liivin kielen sana alla olevaan kenttään.

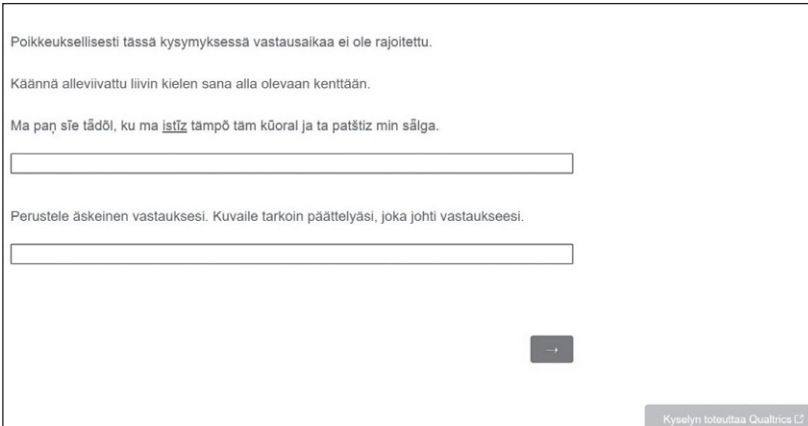
abbönd

→

Kyselyn toteuttaa Qualtrics

KUVA 1. Suomalaisten koe, jossa on käynnissä liivin yksittäisten sanojen osuus

Rajasin vastausajan, jotta koe ei rasittaisi vastaajia. Toisaalta vastausaikaa tarvittiin riittävästi, jotta koehenkilöt ehtisivät miettiä vastauksen ja kirjoittaa sen. Kokeessa oli 45 sellaista kysymystä, jossa vastausaikaa oli rajattomasti. Niiden yhteydessä vastaus piti kuitenkin myös perustella. Kuva 2 esittää tällaista kysymystä.



Poikkeuksellisesti tässä kysymyksessä vastausaikaa ei ole rajoitettu.

Käännä alleiviivattu liivin kielen sana alla olevaan kenttään.

Ma paņ sie tādai, ku ma istiz tāmpō tam kūoral ja ta patšiz min sālga.

Perustele äskeinen vastauksesi. Kuvaile tarkoin päättelyäsi, joka johti vastaukseesi.

→

Kyselyn toteuttaa Qualtrics

KUVA 2. Suomalaisten koe, jossa on käynnissä liivin lauseisuus eikä kysymyksessä ole aikarajaa

Seuraavaksi kuvaan pilottitutkimusta, jossa viron ja suomen taitotasolla 0, A ja B olevat suomalaiset koehenkilöt ja yksi virolainen koehenkilö käänisivät verkkokokeessa sukukielten sanoja yksittäin, osana lausetta ja siten, että nuo lauseet olivat osana kokonaista tekstiä.

4. Esipilotti ja pilotti

Etsin pilottiini koehenkilöitä erilaisin ilmoituksin muun muassa Tuglas-seuran kautta. Koehenkilöt saivat arvioida tasonsa itse esimerkiksi suorittamiensa kielikurssien pohjalta. Tuloksia analysoidessani huomioin, että itsearviointi voi vaikuttaa arvioiden pätevyYTEEN. Suomalaisia koehenkilöitä oli 66, virolaisia 37. Kokeet tehtiin yksitellen verkossa. Kokeiden jälkeen kultakin taitotasolta haastateltiin kaksi koehenkilöä molemmista koeryhmistä.

Haastatteluissa keskusteltiin, miten koe eteni, mitä keinoja kääntämisessä käytettiin ja millaisia haasteita kääntämisessä oli. Erikseen käsiteltiin tehtäviä, joissa koehenkilö vastasi muista poiketen, yleensä oikein tai lähes oikein. Tarkoituksena oli löytää sellaisia kielten yhteisiä piirteitä, jotka vain harvat vastaajat tunnistivat. Ajan säästämiseksi läheskään jokaista tehtävää ei käsitelty. Haastattelu oli luonteeltaan puolistrukturoitu (ks. Hirsjärvi & Hurme 2022: 4.2.3), ja haastattelukysymykset on esitetty liitteessä 1.

4.1. Toteutus

Varsinaisia pilottikokeita edeltävä esipilottikoe oli kesällä 2022. Sen ja pilottien tarkoituksena oli varmistaa, että varsinaiset kokeet onnistuvat teknisesti. Toinen tärkeä syy oli, että alustavien kokeiden avulla pystyin varmistamaan, että varsinaisista kokeista saatavat vastaukset ovat sellaisia, että niiden avulla voin myöhemmin vastata tutkimuskysymyksiini. Esipilotin teki äidinkielenään suomea puhuva 31-vuotias mies, jonka viron taitotasoa oli 0 ja korkein koulutus ylioppilastutkinto. Valvoin kokeen samassa tilassa, ja koe kesti 1 tunnin ja 45 minuuttia. Lopuksi

tehtiin noin 15-minuuttinen haastattelu. Esipilottikokeessa ilmeni muutamia teknisiä pulmia, jotka korjattiin.

Seuraavana oli kokeen varsinainen pilotointi. Suomalaisten koe pilotoitiin joulukuussa 2022. Siihen otti osaa kaksi henkilöä, toinen taitotasolta B2, toinen tasolta A. Näistä pilottikokeista alkaen koehenkilöt tekivät kokeet itsenäisesti verkossa, mikä onnistui teknisesti hyvin. Sen jälkeen heidät tarvittaessa haastateltiin Zoom-videopuheluohjelman välityksellä. Suomalaisten pilottikokeen jälkeen kokeeseen tehtiin vielä muutama tekninen korjaus, minkä jälkeen suomenkielisten testaus alkoi. Vironkielisten kokeen pilotti järjestettiin tammikuussa 2023, ja varsinaiset testaukset alkoivat sen jälkeen.

Esipilotin tulokset pisteytettiin luvussa 3.4 kuvatuin linjauksin. Jokaisen osion enimmäispistemäärä oli 17,000 pistettä. Pisteytyksen luotettavuuden arvioimiseksi lisäksi toinen, pisteytykseen perehdyttämäni, henkilö pisteytti kolmen pilottiin osallistuneen vastaukset. Hän on suomalais-ugrilaisen kielitieteen ylioppilas. Hänen äidinkieltensä on suomi ja viron taitotaso C2.

Toinen pisteyttäjä pisteytti ensin ensimmäisen suomalaisvastaajan liivin vastaukset ja osan viron osien vastauksista. Virolaisen koehenkilön vastauksista avustaja pisteytti vepsän vastaukset ja osan suomen osien vastauksista. Pisteytysten vertailu vastauksittain osoitti, että pisteytys oli noin 89,7-prosenttisesti yhteneväistä. Sitten yhtenäistimme pisteytyksen linjaa siten, että sovimme eräiden rajatapauksien pisteyttämisestä. Sitten hän pisteytti loppuun näiden kahden koehenkilön vastaukset ja kaikki toisen suomalaisen koehenkilön vastaukset. Uusien pisteytysten yhdenmukaisuus oli noin 93,5 prosenttia. Kaikkiaan pisteytimme vastauksia noin 92,4-prosenttisesti samalla tavalla. Pisteytysjärjestelmä on tämän perusteella niin luotettava, että sen avulla saatavia tuloksia voidaan käyttää varsinaisten kokeiden vastausten arviointiin. Esipilottiin ja pilottiin osallistuneiden koehenkilöiden koepisteet on kuvattu taulukossa 4, jossa suomalaisten osalta heikoimmat tulokset on alleviivattu, parhaat vahvennettu.

TAULUKKO 4. *Esipilotin ja pilotin koetulokset*

Koehenkilö	Vaihe	Äidin- kieli	Viron taitotaso	Liivin osion pisteet	Viron osion pisteet	Vepsän osion pisteet
Mies, 31 v.	esi- pilotti	suomi	0	<u>3,125</u> <u>5,125</u> <u>5,750</u>	<u>6,500</u> <u>8,375</u> <u>8,500</u>	<u>13,125</u> <u>15,125</u> <u>15,125</u>
Mies, 57 v.	pilotti	suomi	B2	6,625 11,125 12,375	15,125 15,375 16,625	15,750 16,625 16,625
Nainen, 68 v.	pilotti	suomi	A	6,500 12,125 12,625	16,375 16,250 16,250	13,750 16,750 16,750
Koehenkilö	Vaihe	Äidin- kieli	Suomen taitotaso	Vepsän osion pisteet	Suomen osion pisteet	Liivin osion pisteet
Nainen, 34 v.	pilotti	viro	A1	9,875 11,750 11,125	11,875 12,250 14,875*	9,625 15,625 17,000

* Ohjelmointivirheen takia tällä vastaajalla tämän osion enimmäispisteet olivat 16,000 pistettä 17,000:n sijaan.

Esipilottiin osallistuneen pistemäärät nousivat kokeen kolmen pääosion välillä selvästi, kun taas viroa osaavilla suomalaisilla tulostaso oli samanlainen viro ja vepsän eri osioissa. Esipilotin vastaajan pisteet nousivat myös verrattaessa yksittäissanojen kääntämistä ja lauseissa olleiden sanojen kääntämistä. Pisteet eivät kuitenkaan juuri nousseet lause- ja tekstiosien välillä. Kun katsotaan viroa osanneiden suomalaisten tuloksia, nähdään, että he saivat liivin yksittäissanojen osiosta paremmat pisteet kuin viroa taitamaton suomalainen. Samoin näiden kahden suomalaisen pisteet nousivat myös liivin yksittäissanoista lausetasolle siirryttäessä, aivan kuten esipilottikokeeseenkin osallistuneella. Tuloksista myös havaitaan, että vaikka esipilottiin osallistunut sai alhaisemmat pisteet liivin lause- ja tekstiosista, kenenkään pistemäärät eivät juuri nousseet lause- ja tekstiosien välillä.

Joulukuun toisen pilottikokeen vepsän osion alussa koehenkilö vastasi muutamaan kysymykseen täysin väärin yksittäisiä sanoja kääntäessään. Näihin kysymyksiin hän vastasi viron osiossa antamallaan käännöksillä. Viron osalta hän saavutti lähes täydet pisteet joka osiossa. Kokeen loppukyselyssä hän kertoi: ”Kun teksti oli sama kaikissa kieliversioissa, se alkoi tuntua tutulta ja jos itse sana näytti hyvinkin vieraalta, saattoi myös muistella muita kieliversioita.” Hän siis lienee arvaillut vastauksia vepsän yksittäisten sanojen osiossa edellisten vastausten perusteella. Se johti selvästi heikompiin tuloksiin kuin vepsän kokeen muissa osioissa. Kokonaisuuden kannalta tapaus oli hyödyksi, sillä arvelin jo koetta suunniteltaessa, että näin voi tapahtua.

Virolaisten pilottikoe järjestettiin henkilölle, jonka suomen taitotaso oli A1. Hän osasi kahdeksaa vierasta kieltä, joiden joukossa oli toinenkin suomalais-ugrilainen kieli, unkari, taitotasoltaan A1. Hänen suomen kielen lauseosiossaan oli ohjelmointivirheen vuoksi vain 16 kysymystä, eli sen enimmäispisteet olivat poikkeuksellisesti 16,000 pistettä, eivät 17,000. Korjasin ohjelmointiviat.

4.2. Esipilotin ja pilotin tuloksista

Pilottikokeiden perusteella kokeen suunnittelua ja asetelmaa voi pitää onnistuneena. Näin pienistä otoksista ei kannata laskea tilastollista merkittävyyttä, eli koetulokset edustavat yksittäisiä vastaajia, eivät heidän edustamiaan ryhmiä, esimerkiksi kielitaidon osalta. Kokeiden tuloksiin ei siis vielä tule kiinnittää suurta huomiota, vaan siihen, että pilottikokeiden perusteella koe sopii haluttujen erojen esiin saamiseen.

Näin ollen kokeen avulla voidaan myöhemmin saada tuloksia, joilla voidaan tutkia, toteutuuko hypoteesini, jonka mukaan sukukielen taidosta on etua niin eteläryhmään kuuluvaa liiviä kuin pohjoisryhmään kuuluvaa vepsää käännettäessä (ks. koesanojen eroja taulukoista 1–2). Tähän viitannee esimerkiksi se, että esipilotin vastaajan pisteet nousivat kokeen kolmen pääosion välillä selvästi, mutta viroa opiskelleiden pisteet nousivat vain liivin osiosta viroon siirryttäessä. Se voi osoittaa, että

liivin yksittäissanat olivat heille lähes yhtä vaikeita kuin viroa taitamattomalle suomalaisellekin, mutta muissa osioissa saavutetut verrattain korkeat pistemäärät johtuivat kielitaidon tuomasta edusta.

Kaikkien kolmen suomalaisvastaajan pistemäärät olivat vepsän osiossa samaa luokkaa. Varsinaisia koetuloksia analysoitaessa voitaneenkin saada vahvistus sille, että viroa taitotasosta ja kokeen kuormittavuudesta huolimatta altistus eri lähisukukielille kokeen aikana johtaa lopulta viimeisessä osiossa samantasoiseen ymmärtämiseen. Hypoteesini on, että käännettävien sanojen ympärillä olevien tekstin sisäisten vihjeiden ansiosta kääntäminen helpottuu, kuten esimerkiksi Muikku-Werner (2014: 115–117, 2015: 209–211) on osoittanut. Tätä täydentää olettamukseni, että lauseosioista tekstiosioon siirryttäessä tapahtuu kuitenkin kylläntyminen ja tekstin määrän kasvaminen käännettävän sanan ympärillä ei enää juuri tuo helpotusta kääntämiseen.

Pilottikokeita voi pitää onnistuneena siinäkin mielessä, että ne osoittivat, että kokeella voidaan kerätä aineistoa, jolla voidaan tutkia ymmärtämisestä itämerensuomalaisten kielten alaryhmien välillä. Verrattaessa suomea taitavan virolaisvastaajan tuloksia kahteen viroa taitavaan suomalaiseen tulokset ovat erilaisia. Hän sai yksittäisiä vepsän sanoja kääntäessään selvästi korkeammat pisteet kuin nämä kaksi suomalaista liivin yksittäisten sanojen osiosta. Vepsän lause- ja tekstiosioissa hän kuitenkin sai samaa luokkaa olevat pisteet kuin suomalaiset liivin lause- ja tekstiosioissa. Suomen osiosta hän suoriutui epätasaisemmin kuin suomalaiset viroa osiosta ja sai vasta tekstiosioista samaa luokkaa olevat pisteet kuin suomalaiset viroa osiossa. Liivin osion tulokset olivat myös erilaisia kuin suomalaisten vepsän osion tulokset, sillä suomalaiset saivat vepsässä lähes täydet pisteet joka osiossa.

Verrattaessa suomalaisten vepsän tuloksia virolaisen liivin osioon voidaan tarkastella, miten koe toimii näillä ryhmillä lähintä sukukieltä käännettäessä. Virolaisvastaaja sai liivin kokeessa lause- ja tekstiosiossa samanlaisia pistemääriä kuin suomalaiset vepsän lause- ja tekstiosiossa. Yksittäisten liivin sanojen kääntäminen oli virolaiselle kuitenkin vaikeampaa kuin suomalaisille vepsän yksittäissanon kääntäminen. Liivin

yksittäissanojen osuudessa hän sai jopa heikommat pisteet kuin kääntäessään vepsän ja suomen yksittäissanoja. Hän oli silti ainoa, joka sai jossain osiossa täydet pisteet käännettyään liivin tekstiosiossa kaikki sananmuodot täysin oikein.

5. Päätäntö

Pilottikokeiden perusteella kokeen suunnittelua ja sen asetelmaa voidaan pitää onnistuneina. Kokeen avulla voin tarkkailla tutkimuskysymyksiäni eli voin selvittää, miten vastineiden havaitseminen ja ymmärtäminen muuttuvat, kun vastineet esiintyvät yksittäin, lauseissa tai tekstien osana; miten suomalaisten ja virolaisten kielitaito vaikuttaa eri yhteyksissä esiintyvien vastineiden ymmärtämiseen; miltä osin lähisukukielten ymmärtäminen on symmetristä ja miltä osin epäsymmetristä; ja millaista on itämerensuomalaisten kielten pohjois- ja eteläryhmien sisäinen ja välinen keskinäinen ymmärrettävyys. Pilotissa ei ollut merkittäviä teknisiä pulmia, eivätkä koehenkilöt pitäneet koetta liian vaikeana tai liian kuormittavana. Testin myöhemmin julkaistavat varsinaiset tulokset näyttävät, toimiiko myös näille kielille altistuminen eli itse koeaineisto kääntämisen apuna. Jo pilottikokeissa koehenkilöt totesivat, että he havaitsivat kokeessa käännettävien sanojen olevan toistensa vastineita. Tosin yhtä vastaajaa liika koeaineistoon nojaaminen johti hetkellisesti harhaan.

Pisteytysjärjestelmä ei ole altis arvioijien väliselle vaihtelulle, eli vastaajien pisteitä voi perustellusti vertailla. Pisteytysjärjestelmä tuo uuden työkalun van Bezooijen ja Gooskensin menetelmän ja muokkausetäisyyden rinnalle, sillä järjestelmä huomioi tarkasti ymmärtämisen eri asteet esimerkiksi muodon ja merkityksen havaitsemisen kannalta. Järjestelmä ei ole niin väljä ja subjektiivinen kuten van Bezooijen ja Gooskensin menetelmä tai niin mekaaninen ja kankea kuin muokkausetäisyys. Vastausten perusteluista tulee toivottavasti käymään ilmi myös se, huomaa-atko vastaajat koesanojen yhteisen alkuperän.

Keskeisimpiä havaintojani on, että kokeen avulla voin vastata tutkimuskysymyksiini ja ottaa samalla huomioon koehenkilöiden äidinkielen

ja kielitaidon. Siihen viittaavat pilottien piste-erot viroa taitavien suomalaisten ja viroa taitamattoman suomalaisen välillä. Myös erot koemenestyksessä viroa taitavien suomalaisten liivin ja vepsän osioissa suomea taitavan virolaisen vepsän ja liivin osioihin nähden viittaavat kokeen suunnittelun onnistumiseen. Tutkimuskysymyksiin saadaan vastauksia, kun seuraavaksi analysoin 66 suomalaisen ja 37 virolaisen koehenkilön vastaukset perusteluineen sekä heidän haastattelunsa.

Kiitokset

Esitän kiitokset koetekstistä huomioita tehneille; käänöksessä auttaneille professori Riho Grünthalille, filosofian tohtori Nina Zaitsevalle ja professori Valt Ernštreitille; avustajilleni Hanna-Mari Räsäselle, Tiina Kloosterille ja Elsa Mensalolle; koehenkilöille; Tuglas-seuralle; sekä FM Bogáta Timárilille ja FM Netta Lehtikankaalle.

Lähteet

- Alekseeva, Irma 2016. Vienankarjalan ymmärtäminen suomen kielen pohjalta. Maisterintutkielma. Jyväskylän yliopiston kielten laitos.
- van Bezooijen, Reneé, Charlotte Gooskens 2007. Interlingual text comprehension: Linguistic and extralinguistic determinants. – Jan D. ten Thije, Ludger Zeevaert (eds.). *Receptive Multilingualism: Linguistic Analyses, Language Policies and Didactic Concepts*. Hamburg Studies on Multilingualism 6. Amsterdam–Philadelphia: John Benjamins Publishing Company, 249–264. <https://doi.org/10.1075/hsm.6.17bez>
- Bolotova, Natalia, Nikolay Rakin 2007. Lähtökielenä sukukieli ja venäjä: Venäjän suomalais-ugrialaisten kielten puhujien kokemuksia suomen ja unkarin oppimisesta. – Pirkko Muikku-Werner, Ossi Kokko, Hannu Remes (toim.). *VIRSU III: Suomalais-ugrilaisia kohdekieliä ja kontakteja*. *Studies in Languages* 42. Joensuu: Joensuun yliopisto.
- Borin, Lars 2013. Introduction: The why and how measuring linguistic differences. – Lars Borin, Anju Saxena (eds.). *Approaches to Measuring Linguistic Differences*. *Trends in Linguistics. Studies and Monographs* 265. Berlin: Walter de Gruyter, 1–28. <https://doi.org/10.1515/9783110305258>
- Grünthal, Riho 2015. Livonian at the crossroads of language contacts. – Santeri Junttila (ed.). *Contacts Between the Baltic and Finnic Languages*. *Uralica Helsingiensia* 7. Helsinki: Helsingin yliopisto, 12–67.

- Heeringa, Wilbert, Jelena Golubovic, Charlotte Gooskens, Anja Schüppert, Femke Swarte, Stefanie Voigt 2014. Lexical and orthographic distances between Germanic, Romance and Slavic languages and their relationship to geographic distance. – Charlotte Gooskens, Reneé van Bezooijen (eds.). *Phonetics in Europe: Perception and Production*. Frankfurt am Main, etc.: Peter Lang, 99–137.
- Heikkinen, Vesa, Outi Lehtinen, Mikko Lounela 2001. Kuvia kirjoitetusta suomesta. – *Kielikello* 34. <https://www.kielikello.fi/-/kuvia-kirjoitetusta-suomesta> (6.9.2023).
- Hirsjärvi, Sirkka, Helena Hurme 2022. *Tutkimushaastattelu: Teemahaastattelun teoria ja käytäntö*. Helsinki: Gaudeamus.
- Härmävaara, Hanna-Ilona 2022. *Building the Bridge in Interaction: Receptive Multilingualism Among Finnish and Estonian Speakers*. Väitöskirja. Helsingin yliopisto. <http://urn.fi/URN:ISBN:978-951-51-7921-0> (27.7.2023).
- Härmävaara, Hanna-Ilona, Charlotte Gooskens 2019. Mutual intelligibility of Finnish and Estonian vocabulary. – *Lähivördlusi. Lähivertailuja* 29, 13–56. <https://doi.org/10.5128/LV29.01>
- Kaivapalu, Annekatrin 2004. Kui sarnane on sarnane? Eesti ja soome mitmusevormide psühholingvistilisest reaalsusest. – Helena Sulkala, Heli Laanekask (toim.). *VIRSU II: Suomi ja viro kohdekielinä*. *Lähivördlusi. Lähivertailuja* 15. Oulun yliopiston suomen ja saamen kielen ja logopedian laitoksen julkaisuja 24. Oulu: Oulun Yliopisto, 62–71.
- Kaivapalu, Annekatrin 2005. Lähdekieli kielenoppimisen apuna. *Jyväskylä Studies in Humanities* 44. Jyväskylä: Jyväskylän yliopisto.
- Kaivapalu, Annekatrin 2009. Vironkielisen suomenoppijan äidinkieli – ongelma haaste vai voimavara? – *Virittäjä* 103, 382–401.
- Kaivapalu, Annekatrin 2013. Kielten järjestelmien vertailusta kieltenvälisen vaikutuksen tutkimukseen. – Leena Kolehmainen, Matti Miestamo, Taru Nordlund (toim.). *Kielten vertailun metodiikka. Suomalaisen Kirjallisuuden Seuran toimituksia* 1387. Helsinki: SKS, 293–323.
- Kaivapalu, Annekatrin 2015. Eesti ja soome keele vastastikune mõistmine üksik-sõna- ja tekstitasandil: lingvistilised tegurid, mõistmisprotsess ja sümmeetria. – *Eesti rakenduslingvistika ühingu aastaraamat* 11, 55–74. <https://doi.org/10.5128/ERYa11.04>
- Kaivapalu, Annekatrin, Pille Eslon 2011. Onko lähisukukielen vaikutus suomen ja viron omaksumiseen symmetristä? Korpuspohjaisen tutkimuksen tuloksia ja haasteita. – *Lähivördlusi. Lähivertailuja* 21, 132–153. <https://doi.org/10.5128/LV21.06>

- Kaivapalu, Annekatrin, Maisa Martin 2014. Measuring perceptions of cross-linguistic similarity between closely related languages: Finnish and Estonian noun morphology as a testing ground. – Heli Paulasto, Lea Meriläinen, Helka Riionheimo, Maria Kok (eds.). *Language Contacts at the Crossroads of Disciplines*. Cambridge: Cambridge Scholars Publishing, 283–318.
- Kaivapalu, Annekatrin, Maisa Martin 2015. Kuidas mõjutab emakeele keeltevahelise sarnasuse tunnetamist. – *Emakeele Seltsi aastaraamat* 60, 74–100. <https://doi.org/10.3176/esa60.04>
- Kaivapalu, Annekatrin, Maisa Martin 2017. Perceived similarity between related languages: Strings of letters or morphological units? – *Nordic Journal of Linguistics* 40, 149–174. <https://doi.org/10.1017/S0332586517000142>
- Kaivapalu, Annekatrin, Pirkko Muikku-Werner 2010. Reseptiivinen monikielisyys: miten suomenkielinen oppija ymmärtää viroa äidinkielsä pohjalta? – *Lähivõrdlusi. Lähivertailuja* 20, 68–97. <https://doi.org/10.5128/LV20.03>
- Kallio, Petri 2007. Kantasuomen konsonanttihistoriaa. – Jussi Ylikoski, Ante Aikio (toim.). *Sámit, sánit, sátnehámit: riepmočála Pekka Sammallahtii miessemánu 21. beaivve 2007*. Suomalais-Ugrilaisen Seuran toimituksia 253. Helsinki: SUS, 229–249.
- Kallio, Petri 2024. Suomenlahden eteläpuolisia itämerensuomalaisia sekamurteita. – Sofia Björklöf, Santra Jantunen, Santeri Junntila, Juha Kuokkala, Saarni Laitinen, Annika Pasanen (toim.). *Itämeren kieliapajilta Volgan verkoille: Pühendusteos Riho Grünthalile 22. mail 2024*. Suomalais-Ugrilaisen Seuran toimituksia 278. Helsinki: SUS, 305–320.
- Kirjoitetun suomen kielen sanojen taajuuksia. Kotimaisten kielten keskus. <http://kaino.kotus.fi/sanat/taajuuslista/parole.php> (8.6.2020).
- Lehtinen, Tapani 2008. Sukukielet syynissä: Uralilaiset kielet ja niiden tutkimus. Tiina Onikki-Rantajääskö, Mari Siirainen (toim.). *Kieltä kohti*. Helsinki: Otava, 233–258.
- Mantila, Harri 2000. Meänkieli, yksi Ruotsin vähemmistökielistä. – *Kielikello* 33, 9–12.
- Massinen, Henna 2022. Rajakarjalaismurteiden suomalaistuminen yleisgeminatian ja itämurteiden erikoisgeminatian valossa. *Publications of the University of Eastern Finland. Dissertations in Education, Humanities, and Theology* 186. Joensuu: Itä-Suomen yliopisto. <http://urn.fi/URN:ISBN:978-952-61-4549-5>
- Muikku-Werner, Pirkko 2014. Koteksti ja viron ymmärtäminen lähisuku-kielen pohjalta. – *Lähivõrdlusi. Lähivertailuja* 24, 100–124. <https://doi.org/10.5128/LV24.04>

- Muikku-Werner, Pirkko 2015. Tekstin semanttiset sidokset ja lähisukukielen ymmärrettävyys. – *Lähivõrdlusi. Lähivertailuja* 25, 191–216. <https://doi.org/10.5128/LV25.08>
- Muikku-Werner, Pirkko 2016. Monitasoinen samankaltaisuus: virolaiset ja suomenkielisen tekstin ymmärtäminen. – *Lähivõrdlusi. Lähivertailuja* 26, 311–338. <https://doi.org/10.5128/LV26.10>
- Muikku-Werner, Pirkko, Maria Heinonen 2012. Lumesadu – 'tarina' vai 'lumikasa' vai ei kumpikaan? Suomalaiset lukiolaiset viron sanoja tunnistamassa. – *Lähivõrdlusi. Lähivertailuja* 22, 157–187. <https://doi.org/10.5128/LV22.06>
- Muikku-Werner, Pirkko, Jarmo Harri Jantunen 2020. Understanding Estonian phraseological units on the basis of Finnish: Contributing and misleading factors. – *Lähivõrdlusi. Lähivertailuja* 30, 155–186. <https://doi.org/10.5128/LV30.04>
- Muikku-Werner, Pirkko, Helka Riionheimo 2018. Livvinkarjalaisen tekstin ymmärtäminen suomen pohjalta. – *Lähivõrdlusi. Lähivertailuja* 28, 285–315. <https://doi.org/10.5128/LV28.09>
- Mustonen, Ari 2014. Pääosin ohjaamaton sanaston poiminta rakenteettomasta tekstistä. Diplomityö. Oulun yliopisto. <http://jultika.oulu.fi/files/nbn-fioulu-201404081248.pdf> (10.9.2023).
- Mutual Intelligibility of Closely Related Languages (MICReLa). Research projects. <https://www.let.rug.nl/gooskens/project/?p=projects> (5.7.2024).
- O'Rourke, Patrick, Karl Pajusalu 2016. Livonian features in Estonian dialects. – *ESUKA/JEFUL* 7, 67–85. <https://doi.org/10.12697/jeful.2016.7.1.03>
- Paajanen, Ilona, Pirkko Muikku-Werner 2012. Tee on kitsas – onko 'tee kitkerää' vai oletteko 'te saita'? Suomalaiset opiskelijat viroa ymmärtämässä. – *Lähivõrdlusi. Lähivertailuja* 22, 219–258. <https://doi.org/10.5128/LV22.08>
- Palokangas, Aleksi 2018. Viron yleisten sananmuotojen ymmärtäminen. Masterintutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:ju-201812195242>
- Sammallahti, Pekka 1977. Suomalaisten esihistorian kysymyksiä. – *Virittäjä* 81, 119–136.
- Tartu Ülikooli arvutilingvistika uurimisrühm 2019a. Sõnaliikide sagedusloend ning käändsõna grammatiliste kategooriate sagedusloendid Tasakaalus korpuse põhjal. – <https://www.cl.ut.ee/ressursid/gram-kat/> (7.9.2023).
- Tartu Ülikooli arvutilingvistika uurimisrühm 2019b. Tabel 2. Sõnaliikide sagedused. – <https://www.cl.ut.ee/ressursid/gram-kat/tabel2/> (7.9.2023).
- Tartu Ülikooli arvutilingvistika uurimisrühm 2024. Tasakaalus korpuse sõnavormid sageduse järjekorras. – <https://www.cl.ut.ee/ressursid/sagedused1/index.php?lang=et> (9.5.2024).

Liite. Puolistrukturoidun haastattelun runko

1. Millaista oli tehdä koetta? Sujuiko koe ongelmitta?
2. Millä laitteella teit kokeen, älypuhelin, tietokone vai tabletti?
3. Kauanko kokeen tekemiseen meni aikaa?
4. Oliko kokeessa jotain epäjohdonmukaista?
5. Olivatko ohjeet selkeät?
- (6. Miten muuttaisit koetta? Miksi tekisit kuvailemiasi muutoksia?)
7. Käytitkö jotain tiettyä lähestymistapaa tai strategiaa kääntämisessä? Miten hyödynsit sitä?
8. Vaihteliko lähestymistapasi kielten välillä?
9. Hyödynsitkö käännettävän sanan ympärillä olevia sanoja tai joitain muita seikkoja kääntäessäsi? Kuvaile käyttämiäsi menetelmiä.
10. Millaiset tehtävät koit vaikeimmiksi? Sellaiset, joissa käännettiin yksittäisiä sanoja, sanoja lauseissa vai sanoja tekstikatkelmissa?
11. Vaihteliko tämä kielittäin?
12. Mikä on suhteesi Viroon, viron kieleen ja virolaisiin? (Virolaisilta: Mikä on suhteesi Suomeen, suomen kieleen ja suomalaisiin?)
13. Mitä tiedät vepsän kielestä ja liivin kielestä tai ylipäätään näistä kokeen kolmesta kielestä?
14. Käytitkö suomea apuna kääntämisessä? (Käytitkö viroa apuna kääntämisessä?)

Planning and novelty values of a test measuring mutual intelligibility of Baltic Finnic languages

ALEKSI PALOKANGAS

University of Helsinki

Mutual intelligibility of Baltic Finnic languages can be measured and observed in multiple ways. This article describes the development of a new kind of language test designed for informants whose mother tongue is Finnish or Estonian. The participating Finns' language proficiency level in Estonian was 0, A or B, and Estonians' level in Finnish was 0, A or B. Compared to previous test types, ways to measure mutual intelligibility, and measuring differences between cognates this test offers new features. The process of test word selection, test design, and scoring method are described phase by phase.

The selection of cognates was conducted with meticulous care, as the test aimed to provide diverse data for analysis. One of the most important purposes of the test was to observe and analyse how these two groups of people understood cognates from each other's language in different contexts. First, the cognates were translated one at a time, then as a part of sentences, and ultimately as a part of short text. Additionally, informants translated Vepsian and Livonian languages.

Another objective in developing this test was to collect data for analysing whether Finns' language skills in Estonian aided their understanding of Livonian and whether Estonians' language skills in Finnish helped them to understand Vepsian. Hence, the informants not only translated cognates shared by all these languages but also those with specific etymological connections. For instance, they considered cognates like 'parta' and 'bard' (both meaning 'beard') – which have a common etymology only in Finnish and Vepsian – as well as 'habe' and 'abbõnd' (also meaning 'beard') – which share a common etymology exclusively between Estonian and Livonian.

Furthermore, this type of test provides an opportunity to analyse how symmetric or asymmetric the level of understanding between Finnish and Estonian is. Therefore, the cognates in the test include morphologically rich word forms

such as Finnish *yhden*, Vepsian *ühten*, Estonian *ühe* ja Livonian *id* which all translate to 'one's', and *pitkä*, *pit'k*, *pikk*, *pitkā* ('tall'), where Estonian *pikk* is a false friend for Finns as it resembles Finnish word *pikku* ('little, small, tiny').

The piloting of the test and its results have also been described in this article, alongside the development process of the test's scoring system. The test's scoring method is thoroughly explained, as it is nuanced rather than binary. The pilot's results substantiate that the development, structure and scoring method of the test were successful. The test fulfills its purpose and objectives as it is capable of highlighting differences among informants with varying backgrounds, such as different mother tongues and language proficiency levels.

Keywords: understanding of a second language; mutual intelligibility; language test; receptive multilingualism; Finnish; Estonian; Vepsian; Livonian

Aleksi Palokangas

P.O. Box 4, 00014 University of Helsinki, Finland
aleksi.palokangas@helsinki.fi