

Onko lähisukukielen vaikutus suomen ja viron omaksumiseen symmetristä? Korpuspohjaisen tutkimuksen tuloksia ja haasteita

ANNEKATRIN KAIVAPALU

Jyväskylän yliopisto, Tallinnan yliopisto

PILLE ESLON

Tallinnan yliopisto

Tiivistelmä. Artikkeliki käsittelee kahden lähisukukielen, viron ja suomen, vaikutusta toistensa omaksumiseen oppijanviron ja oppijansuomen korpusten pohjalta. Artikkelissa keskitytään lähisukukielen morfologiseen vaikutukseen viron ja suomen kielen taitotasolla B1 ja B2. Tutkimuksen kohteena on lähdekielen vaikutuksen mahdollinen symmetrisyys: oletuksena on, että jos lähde- ja kohdekielen järjestelmien välillä on symmetriasuhde, todellinen samanlaisuus, niin myös lähdekielen vaikutus toimii symmetrisesti molempiin suuntiin. Kieltenvälinen todellinen samankaltaisuus nähdään jatkumona maksimaalisesta konvergenssista maksimaaliseen divergenssiin.

Tutkimus osoittaa, että lähdekielen positiivinen morfologinen vaikutus ilmenee symmetrisesti vironkielisten suomenoppijoiden ja suomenkielisten vironoppijoiden kirjallisissa tuottamisprosesseissa: oppijat käyttävät eniten suomessa ja virossa samalla tavalla taipuvia muotoja. Lähde- ja kohdekielen taivutusmallien konvergenssi on vironkielisille suomenoppijoille läpinäkyvämpi kuin suomenkielisille vironoppijoille, jotka tarvitsevat taivutusmallien

konvergenssin havaitsemisen tueksi enemmän semanttista ja fonologista konvergenssia.

Avainsanat: toisen kielen oppiminen; kieltenvälinen vaikutus; korpuspohjainen tutkimus; lähisukukielet; konvergenssi; divergenssi; morfologia; nominitaivutus; suomi; viro

1. Taustaa

Tämä pilottitutkimus on osa Tallinan yliopiston viron kielen ja kulttuurin instituutin hanketta *Kieltenvälinen vaikutus ja toisen kielen omaksuminen: korpuspohjainen tutkimus* (2010–2013) sekä Jyväskylän yliopiston soveltavan kielentutkimuksen keskuksen ja Tallinnan yliopiston viron kielen ja kulttuurin instituutin yhteishanketta *Symmetry of the cross-linguistic influence in the acquisition of closely related languages* (2010–2013), jotka perustuvat viron ja suomen oppijankielen korpusten rinnakkaiseen käyttöön. Hankkeiden tavoitteena on

- 1) tarkastella oppijankielen korpusten pohjalta lähdekielen morfologista, morfosyntaktista ja leksikaalista vaikutusta suomen ja viron omaksumiseen sekä selvittää lähisukukielen ja ei-sukukielen vaikutuksen eroja,
- 2) tarkastella lähdekielen vaikutuksen ja kielitaitotason suhdetta,
- 3) selvittää, millaiset kielen omaksumiseen ja prosessointiin vaikuttavat tekijät toimivat yhdessä lähdekielen vaikutuksen kanssa, sekä
- 4) selvittää, millaisia ovat erikielisten oppijoiden prosessointistrategiat.

Artikkelissamme keskitymme lähisukukielen morfologiseen vaikutukseen viron ja suomen kielen taitotasoilla B1 ja B2. Tavoitteemme on myös menetelmällinen: tarkoituksenamme on testata sekä korpusaineiston että käyttämämme metodin sopivuutta lähdekielen vaikutuksen tutkimiseen.

Viron- ja suomenoppijoiden ensikielten vaikutus oppimiseen on ollut ja on pitkälti edelleenkin virheanalyytisessä tutkimuksessa satun-

naisesti esitettyjen arvailujen varassa. Lähdekielen vaikutus on nähty ja nähdään usein ensisijaisesti negatiivisena ilmiönä, virheiden aiheuttajana. Lähdekielen negatiivinen vaikutus eli interferenssi, sen pitkstä tutkimusperinteestä ja läpinäkyvyydestä huolimatta, on lähdekielen positiiviseen vaikutukseen verrattuna kuitenkin toissijaista. Opetellessaan uutta kieltä oppija joko tietoisesti tai tiedostamattaan hyödyntää aikaisempaa tietoa ja taitoa äidinkielestään ja/tai muista jo tuntemistaan kielistä. Kieltä opetteleva soveltaa kohdekieleen maailmantietoa, leksikaalis-kieliopillisia rakenteita ja ilmaisustrategioita, joita hän ennestään hallitsee. Kielten lähisukulaisuus ja rakenteellinen samankaltaisuus helpottavat varsinkin kohdekielen ymmärtämistä mutta myös sen tuottamista. Sekä vironkieliset suomenoppijat että suomenkieliset vironoppijat tukeutuvat siis luonnollisesti omiin äidinkieliinsä viroon ja suomeen, jotka varsinkin kieliopilliselta rakenteeltaan mutta myös leksikaalisesti ja funktionaalisesti ovat melko samanlaisia, joiltakin osin jopa melkein identtisiä. Tässä tutkimuksessa tarkastelun kohteena onkin ensisijaisesti lähdekielen positiivinen vaikutus, joka sekä teoreettisesti että metodologisesti on pidettävä selvästi erillään lähdekielen negatiivisesta vaikutuksesta: positiivinen ja negatiivinen lähdekielen vaikutus ilmenevät kohdekielen prosessoinnissa kielen eri omaksumisvaiheissa eri tavalla.

2. Matkalla kohti lähdekielen vaikutuksen päätavoitetta

Lähdekielen vaikutusta kohdekielen omaksumiseen on toisen kielen omaksumisen (SLA) tutkimuksessa pidetty jopa tärkeimpänä kohdekielen omaksumista muokkaavana tekijänä tai ainakin yhtenä erittäin tärkeänä tekijänä muiden joukossa. Alan pitkässä tutkimushistoriassa lähdekielen vaikutusta on käsitelty kieliparikohtaisena yksisuuntaisena ilmiönä tai vertailtu keskenään eri lähdekielten vaikutusta saman kohdekielen omaksumiseen. Tässä artikkelissa lähestymme lähdekielen vaikutusta uudesta näkökulmasta: oletamme, että lähisukukielen vaikutus toisen lähisukukielen omaksumiseen on symmetristä. Etsimme vastausta kysymykseen, ilmeneekö viron ja suomen kielijärjestelmien mahdolli-

nen symmetria symmetriana myös vironkielisten suomenoppijoiden ja suomenkielisten vironoppijoiden tuottamisprosessissa. Tavoitteenamme on siis selvittää, korreloivatko keskenään kielijärjestelmien todellinen ja oppijoiden oletama samanlaisuus, toisin sanoen luoko kielijärjestelmien samanlaisuuden symmetria pohjan myös lähdekielen vaikutuksen symmetrisyydelle.

Lähdekielen vaikutuksen tutkimisen lopullisena tavoitteena on selvittää, miten ihmisen osaamat kielet, sekä äidinkieli että muut myöhemmin opitut tai omaksutut kielet, toimivat vuorovaikutteisesti aivoissa (Jarvis & Pavlenko 2008: 111). Lähdekielen vaikutuksen tutkimisen perinteinen paradigma on rajoittunut vain ilmiön identifiointiin ja kvantifiointiin, kun taas nykyinen tutkimus tähtää sen lisäksi myös lähdekielen vaikutuksen syiden ja rajoitusten selvittämiseen, teoreettisten selitysten kehittämiseen ja aivotoiminnan kuvaukseen. Päämäärän saavuttamiseen tarvitaan empiiristä tutkimustietoa, teoreettisia perusteluja sekä aikaisempaan tutkimukseen pohjautuvien ja myös uusien hypoteesien testausta. (Jarvis 2010.)

Kansainvälinen empiirinen lähdekielen vaikutuksen tutkimus keskittyy yhä enemmän kieltenväliseen samanlaisuuteen pohjautuvaan lähdekielen positiiviseen vaikutukseen, ja tämä suuntaus on laajenemassa muun muassa sekä kognitiivisen aivotutkimuksen että neuropsykologian puolelle. Lähdekielen vaikutuksen kognitiivisen tutkimuksen yhtenä tavoitteena on selvittää, millaisia samanlaisuuksia oppijat pystyvät hyödyntämään suhteessa työmuistiin. Esimerkiksi onko joitakin samanlaisuuksia, jotka ovat tältä osin muita kriittisempiä tai onko olemassa tiettyä samanlaisuuden ”kynnystä”, minimaalista kieltenvälistä samanlaisuutta, joka vaikuttaa positiivisesti työmuistiin. Kohdekielen ymmärtämisen tutkimuksessa on näyttöä siitä, että lähde- ja kohdekielen väliset samanlaisuudet määräävät paljolti sen, millaiset aivoalueet ja miten vahvasti ne aktivoituvat virkkeen ymmärtämistehtävien aikana (Jeong ym. 2007). ERP (engl. *event-related potential*) -efektien positiivinen siirtovaikutus ilmenee myös prosessointitehtävien tuloksissa, kun lähde- ja kohdekielen rakenteet, esim. aikamuodot, sukupuolen ja luvun

ilmaiseminen, kongruoivat, kun taas rakenteiden inkongruenssi estää useissa tapauksissa positiivisen vaikutuksen (Kotz 2009; Tokowicz & MacWhinney 2005).

Oppijankielen korpusten avulla lähdekielen vaikutuksen tutkiminen yhtenä oppijankielen universaalina piirteenä on tullut yhä suosittumaksi (Jantunen 2008: 68). Isojen indoeurooppalaisten oppijankielten korpusten lisäksi on meneillään typologisesti erilaisten kielten, kuten esim. oppijanviron ja oppijansuomen, korpusten kehittäminen ja hyödyntäminen erilaisissa tutkimushankkeissa¹. Laajat sähköiset tietokannat ja kieliteknologiset ratkaisut mahdollistavat tutkimustulosten paremman yleistettävyyden myös lähdekielen vaikutuksen osalta. Tämä on ilmeisesti oppijankielen korpustutkimuksen ja toisen kielen omaksu- misen tutkimuksen lupaavin kosketuspinta (Granger & Meunier 2010).

Toisaalta lähdekielen vaikutuksen korpuspohjainen tutkiminen on monelta osin rajoittunutta (Kaivapalu 2008: 101–112). Sekä perinteinen että korpuspohjainen oppijankielen tutkimus on ollut pääasiassa virheanalyysikeskeistä, mikä on nostanut huomion kohteeksi lähdekielen negatiivisen vaikutuksen. Korpuspohjainen lähdekielen positiivisen vaikutuksen tutkiminen on jäänyt taka-alalle, sillä positiivisen vaikutuksen todistaminen on huomattavasti haasteellisempaa kuin virheanalyysin avulla negatiivisen vaikutuksen todistaminen ja edellyttää erittäin tarkkaa perehtymistä lähde- ja kohdekielen järjestelmiin. Iso haaste on myös kieltenvälisen samanlaisuuden psykolingvistisen todellisuuden eli todellisen, havaitun ja oletetun samanlaisuuden (Ringbom 2007: 7–8) keskinäisten suhteiden selvittely.

Korpuspohjaisessa lähdekielen vaikutuksen tutkimuksessa kohteena on ensisijaisesti oppijoiden tuotoksissa esiintyvien kielen rakenteellisten elementtien siirto kielestä toiseen. Kyseessä on siis enimmäkseen produktin tutkiminen. Tieto prosessien siirrosta on sattumanvaraisempaa, ja siihen pääsee käsiksi vain produktin kautta (Odlin 1989). Korpus-

¹ Kattava yleiskatsaus oppijankielen korpuksista on löydettävissä Esilon & Metslang 2007; suomen ja viron oppijankielen korpuksista ks. Jantunen & Piltunen 2009; koonti meneillään olevista tutkimushankkeista ks. <http://cecl.fltr.ucl.ac.be> (28.2.2011).

aineiston avulla ei myöskään ole mahdollista saada suoranaisesti tietoja oppijoiden kohdekielen tuottamisprosessista, jos ei ole käytettävissä korpuspohjaista tutkimusta täydentäviä introspektiomenetelmiä. Lähdekielen vaikutuksen tutkiminen oppijankielen korpusten pohjalta on siis ensisijaisesti kvantitatiivista.

Vähäiset lähdekielen vaikutusta käsittelevät korpuspohjaiset oppijankielen tutkimukset ovat keskittyneet pääasiassa sanaston oppimisen kysymyksiin, esim. kollokaatioihin oppijanenglannissa (Granger 1998; Nesselhauf 2003; 2005) tai oppijansuomessa (Jantunen 2009; Kallioranta 2009). Jonkin verran on tutkittu myös lähdekielen vaikutusta kohdekielen fonologian omaksumiseen (Rasier & Hiligsmann 2007). Morfologisesti mutkikkaiden oppijankielten, kuten oppijanviron ja oppijansuomen, korpuspohjainen tutkimus tarjoaa erinomaiset mahdollisuudet tarttua lähdekielen vaikutuksen tutkimiseen tällä kansainvälisesti sangen vähän huomiota saaneella alueella.

3. Aineisto, menetelmä ja hypoteesit

Tässä tutkimuksessa keskitymme viron- ja suomenoppijoiden (jatkossa E2- ja S2-oppijoiden) spontaaniin kirjalliseen kielenkäyttöön oppijansuomen ja oppijanviron korpuksissa selvittääksemme, vaikuttaako kirjoitelmissa viron ja suomen morfologisten rakenteiden samanlaisuus oppijoiden taivutusmuotojen valintaan. Tarkoituksenamme on siis selvittää, suosivatko oppijat tuottamisprosessissaan viron ja suomen samanlaisia rakenteita, joiden prosessoinnin suhteellinen helppous saattaa vapauttaa työmuistin kapasiteettia muuhun tuottamistoimintaan. Pohdimme myös sitä, millaiset samanlaisuudet ovat oppijoiden tuottamisprosessin kannalta mahdollisesti ratkaisevia ja mikä voisi olla samanlaisuuden kynnyks, jota eri kielenomaksumisen vaiheissa hyödynnetään.

Lähtökohtanamme on ainakin teoreettisesti kielijärjestelmien vertailun pohjalta selvästi määriteltävissä oleva kieltenvälinen todellinen samanlaisuus (engl. *actual / objective similarity*), joka, jos se on olemassa,

on symmetristä. (Ringbom 2007: 24).² Oletamme myös, että vaikeammin määriteltävissä oleva, usein havaittuun samanlaisuuteen (*perceived similarity*) perustuva ja kohdekielen tuottamisprosessissa olennaisin oletettu samanlaisuus (*assumed similarity*) saattaa (ei-suoranaisesti) ilmetä oppijoiden tuotoksissa.

Kieltenvälisellä vaikutuksella on neljä potentiaalista seurausta tai ilmenemismuotoa (Jarvis 2000: 245–259; 2010: 175–178), jotka ovat käsitettävissä kieltenvälisen vaikutuksen todistetyypeinä tai sen toimimisen edellytyksinä: ryhmänsisäinen homogeenisuus samaa lähdekieltä puhuvien oppijoiden kohdekielen tuotoksissa (*intragroup homogeneity*); ryhmienvälinen heterogeenisuus eri lähdekieliä puhuvien oppijoiden kohdekielen tuotoksissa (*intergroup heterogeneity*); yhteys samaa lähdekieltä puhuvien oppijoiden lähdekielen ja kohdekielen tuotosten välillä (*cross-language congruity*) ja kohdekielen sisäiset kontrastit suhteessa lähdekieleen (*intralingual contrasts*). Ensimmäinen todistetyyppi näyttää, että saman lähdekielen käyttäjät prosessoivat kohdekieltä samalla tavalla; toinen todistaa, että oppijankieli ei ole riippumaton lähdekielistä; kolmas valaisee ja perustelee suhteita lähdekielen vaikutuksen ja sen lähteen välillä; neljäs selvittää lähde- ja kohdekielen ilmiöiden keskinäisiä vastaavuuksia ja sen mukaan sekä positiivisen että negatiivisen lähdekielen vaikutuksen lähtökohtia. Nämä neljä todistetyyppiä yhdessä muodostavat vertailupohjaisen argumentin lähdekielen vaikutuksen esiintymisen puolesta tai sitä vastaan, sillä niiden tutkiminen vaatii kielenkäytön vertailua henkilöiden, ryhmien ja kielten välillä. Tässä tutkimuksessa ovat tarkastelun kohteena ryhmänsisäinen homogeenisuus ja kielenväliset kontrastit.

Pilottitutkimuksemme pohjautuu kahden oppijankielen korpuksen, viron välikielen korpuksen (vir. *Eesti vahekeele korpus*) EVKK:n ja kansainvälisen oppijansuomen korpuksen ICLFI:n, teksteihin.³ Kummankin korpuksen suunnittelussa ja aineistonkeruussa tavoitteena on ollut

² Perusteellisempi suomenkielinen kuvaus kieltenvälisen samanlaisuuden tyypeistä ks. Kaivapalu 2009: 388–389, vironkielinen Kaivapalu 2008: 98–100.

³ <http://evkk.tlu.ee>, <http://www.oulu.fi/oppijankieli> (28.2.2011).

korpuksen keskinäinen maksimaalinen synkronointi sekä kerättävien tekstilajien että oppijoiden taustatietojen osalta, jotta aineistojen vertailtavuus olisi taattu.⁴ Aineistona olemme käyttäneet EVKK:n suomenkielisten oppijoiden ja ICLFI:n vironkielisten oppijoiden osakorpuksia, joihin on koottu oppijoiden kirjoittamia kuvauksia, kertomuksia, arvosteluja, esseitä. Oppijansuomen korpuksessa oli aineiston käsittelyn aikana 530 vironkielisten oppijoiden tekstiä, joiden saneiden kokonaismäärä oli 85 749. Tekstit jakaantuivat eurooppalaisen viitekehysten (2003) kielitaitotasojen kuvausasteikon mukaan seuraavasti (ks. taulukko 1).

TAULUKKO 1. *Vironkielisten oppijoiden aineisto suomen oppijankielen korpuksessa*⁵

Kielitaitotaso	Tekstejä		Saneita	
	n	%	n	%
A1	–	–	–	–
A2	62	12	4 996	6
B1	325	61	39 053	45
B2	120	23	29 046	34
C1	16	3	9 333	11
C2	7	1	3 321	4

On kuitenkin otettava huomioon, että tämä S2-oppijoiden korpuksien kielitaitotasojakauma perustuu vain yhden arvioijan arviointiin, koska oppijansuomen korpuksen arviointi oli tutkimuksemme toteuttamisen aikana kesällä 2010 vasta alkuvaiheessa. Alkuvaiheessa oli silloin myös viron välikielen korpuksen suomenkielisten oppijoiden osakorpuksen kerääminen: tutkimuksen suorittamisen aikana oli oppijanviron

⁴ Viron välikielen korpuksen suunnittelusta ja aineiston keräämisperiaatteista ks. esim. Eslon 2007, suomen oppijankielen korpuksista esim. Jantunen 2007.

⁵ Vironkielisten oppijoiden aineiston kielitaitotasoa koskevista tiedoista kiitämme Tallinnan yliopiston opiskelija Keaty Siiveltia.

korpuksessa suomenkielisten oppijoiden tekstejä 94, siis huomattavasti vähemmän kuin oppijansuomen korpuksessa vironkielisten oppijoiden tekstejä. Korpukselta valittiin ne B1- ja B2-tason tekstit, joiden kielitaitotasosta kaksi arvioijaa olivat samaa mieltä. B1- ja B2- taitotasojen tekstit ovat tarkastelun kohteena ensinnäkin käytännön syistä: niitä esiintyy tavallisesti korpusaineistoissa eniten. Toisaalta B1- ja B2-taso ovat relevantteja myös kielipoliittisesti: Virossa ja Suomessa, kuten useimmissa Euroopan maissa, on B1-taso kansalaisuuden saamisen ehtona ja B2-taso on valtion virkamiehiltä vaadittava kielitaitotaso. Sen lisäksi on B1- ja B2-taitotasojen kommunikatiivinen ero suuri: B2 tason tavoittaminen edellyttää B1-tasoon verrattuna kielitaidon valtavaa kehitystä ihmisen henkilökohtaisen elämän käsittelemiseen riittävästä kielitaidosta yhteiskunnallisessa elämässä pärjäämiseen tarvittavaan kielitaitoon. E2-oppijoiden aineiston mukaisesti valittiin S2-oppijoiden teksteistä sama määrä samantasoisia tekstejä, ja molemmista korpuksista laadittiin pilottikorpukset. Pilottikorpusten aineistoa koskevia tietoja on vertailtu taulukossa 2.

TAULUKKO 2. *E2-oppijoiden ja S2-oppijoiden pilottikorpusten tekstiaineisto*

Pilottikorpus	E2B1	E2B2	S2B1	S2B2
Tekstejä	21	24	21	24
Saneita	3 512	4 607	4 385	5 859
Tekstien pituus	44–589	46–578	63–404	74–549
Tekstien keskipituus	167	192	208	244

Kyseessä on siis pienehkö pilottiaineisto, jonka perusteella tekemämme johtopäätökset eivät ole kovin hyvin yleistettävissä mutta joka oletettavasti tarjoaa silti hyvän pohjan jatkotutkimusta varten ja osoittaa myös mahdollisia ongelmakohtia. Tutkimusmenetelmä on korpuspohjainen (Tognini-Bonelli 2001: 86): tavoitteena oli testata korpusaineiston

avulla aikaisemman läheisen sukukielen vaikutusta käsittelevän tutkimuksen (Kaivapalu 2005) taivutustesteihin pohjautuvia tuloksia. Niiden mukaan vironkieliset oppijat taivuttivat paremmin sanoja, joissa toimii lähdekielen positiivinen vaikutus eli joissa virossa ja suomessa on sama taivutusmalli. Jos taivutusmallin samanlaisuuteen eli lähde- ja kohdekielen morfologiseen konvergenssiin liittyy myös semanttinen ja fonologinen konvergenssi eli sanoilla on virossa ja suomessa sama vartalo ja sama merkitys (esim. *vieras, vilkas, tuote, laine*), taivutustulokset olivat parhaat. Toiseksi parhaat tulokset olivat erivartaloisten, mutta samalla tavalla taipuvien (esim. *ohje, kirje, hidas, patsas*) sanojen ryhmässä, kun taas tilanteessa, jossa viron ja suomen taivutus on erilainen, sama vartalo (esim. *koulu, kirkko, pankki, toinen, kerros, kauppakeskus*) johti oppijat harhaan ja oli siten negatiivisen lähdekielen vaikutuksen aiheuttaja.

Kieltenvälisen samanlaisuuden ja erilaisuuden raja ei siis ole jyrkkä, vaan kyseessä on pikemminkin jatkumo, johon sijoittuvat kaikki mahdolliset kieltenvälisen samanlaisuuden suhteet (ks. myös Tokowicz & MacWhinney 2005; Ringbom 2007; Bochnacker & Rosén 2008; Trofimowich ym. 2008; Jarvis 2010). Taivutustestin tuloksia testattaessa poimittiin pilottikorpuksen teksteistä kaikki nominien taivutusmuodot ja niitä tarkasteltiin suomen ja viron konvergenssin ja divergenssin näkökulmasta. Näistä edellinen on lähdekielen positiivisen vaikutuksen ja jälkimmäinen negatiivisen vaikutuksen edellytys.⁶ Tarkastelun pohjana oli aikaisemman tutkimuksen mukaan neljä lähde- ja kohdekielen konvergenssin ja divergenssin yhdistelmätyyppejä (Kaivapalu 2005: 64–67):

- 1) fonologinen, morfologinen ja semanttinen konvergenssi: sanoilla on virossa ja suomessa sama vartalo, sama taivutus ja sama merkitys (SVST), kuten sm. *lopusta* – vir. *lõpusta*;

⁶ On myös tilanteita, jossa oppijat eivät luota todellisiin kieltenvälisiin samanlaisuuksiin ja siten tekevät virheitä pelkästä uskosta, että kohdekieli ei voi olla samanlainen kuin lähdekieli (ks. esim. Kaivapalu 2005: 97–98, 210–211). Tässä tapauksessa on kyseessä lähdekielen negatiivinen vaikutus tilanteessa, jossa kieltenvälisiä erilaisuutta ei ole.

- 2) fonologinen ja morfologinen konvergenssi, semanttinen divergenssi: sanoilla on virossa ja suomessa eri vartalo ja merkitys, mutta sama taivutusmalli (EVST), kuten sm. *taulusta* – vir. *laulusta*;
- 3) fonologinen ja semanttinen konvergenssi, morfologinen divergenssi: sanoilla on virossa ja suomessa sama vartalo, mutta eri taivutus (SVET), kuten sm. *(ostos)keskuksesta* – vir. *(ostu)keskusest*;
- 4) fonologinen, morfologinen ja semanttinen divergenssi (EVET): sanoilla on virossa ja suomessa sekä eri vartalo että eri taivutusmalli, kuten sm. *tilanteesta* – vir. *olukorrast*.

Tässä artikkelissa esittelemme S2-oppijoiden ja E2-oppijoiden taivutusmuotojen käyttöä yksikön elatiivimuotojen pohjalta. Tällaisen valinnan syitä on useita: 1) suomen ja viron elatiivimuodot ovat rakenteeltaan (genetiivin vartalo, johon agglutinoidaan taivutussuffiksi) samanlaisia viron taivutuspuheen loppuheitosta johtuvaa eroa lukuun ottamatta, erot saattavat johtua vain erityyppisistä vartalovaihteluista; 2) kummasakaan kielessä ei ole yksikön elatiivissa rinnakkaismuotoja, jotka voisivat mahdollisesti kilpailla keskenään oppijoiden tuottamisprosessissa, ja 3) samanrakenteisia sijamuotoja on kummankin kielen taivutusjärjestelmässä muitakin (esim. inessiivi, ablatiivi, allatiivi, adessiivi, ablatiivi, translatiivi), joten samaa menetelmää voi soveltaa laajemminkin aineiston käsittelyyn. Annotoidun oppijankieliaineiston avainsana-analyysin alustavien tulosten (Jantunen 2010) mukaan elatiivimuodot ovat vironkielisten alkeistason⁷ S2-oppijoiden teksteissä aliedustuneita verrattuna natiiviaineistoon, kun taas edistyneiden oppijoiden elatiivinkäyttö ei ole ali- eikä yliedustunutta; toisin sanoin se vastaa siis karkeasti natiiviaineiston elatiivinkäyttöä. Tämä artikkeli on ensimmäinen yritys käsitellä

⁷ Oppijansuomen korpuksen taitotasomääritelmät perustuvat Jantunen (2010) tutkimuksessa opiskelijoiden saaman yliopisto-opetuksen määrään: alkeistason tekstien kirjoittajat ovat saaneet opetusta alle 200 tuntia ja edistyneet yli 400 tuntia. ICLFI:n aineiston arviointi eurooppalaisen viitekehysten taitotasojen kuvausasteikon perusteella on meneillään.

corpusteksteissä esiintyviä elatiivimuotoja ei-homogeenisena ryhmänä viron ja suomen rakenteiden konvergenssin ja divergenssin näkökulmasta.

E2- pilottikorpuksen teksteissä esiintyi yksikön elatiivin muotoja B1-tasolla 23 ja B2-tasolla 47, S2-pilottikorpuksen teksteissä puolestaan B1-tasolla 87 ja B2-tasolla 75. On merkille pantavaa, että kaikki sekä S2-oppijoiden että E2-oppijoiden corpusteksteissä käyttämät elatiivimuodot olivat kohdekielen morfologianormin mukaisia. Tämä tosiasia kysenalistaa taas kerran vain virheanalyysin avulla oppijankielen tutkimisen tarkoituksenmukaisuuden.

Lähtökohtanamme olivat seuraavat hypoteesit:

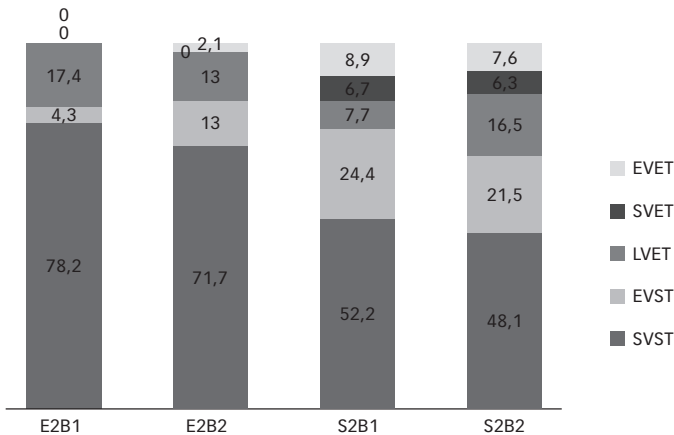
- 1) jos lähde- ja kohdekielen järjestelmien välillä on symmetriasuhde, todellinen samanlaisuus, myös lähdekielen vaikutus toimii symmetrisesti molempiin suuntiin;
- 2) sekä vironkieliset suomenoppijat että suomenkieliset vironoppijat hyödyntävät kohdekielen prosessoinnissa eniten niitä lähde- ja kohdekielessä samanlaisia taivutusmalleja, joiden konvergenssi on maksimissaan, eli fonologinen, morfologinen ja semanttinen.

4. Samanlaisuus jatkumona: alustavia tuloksia

Edellisessä luvussa kuvaamamme viron ja suomen todelliseen samanlaisuuteen ja erilaisuuteen perustuvat sanamuotoryhmat ovat kuvailtavissa jatkumona maksimaalisesta konvergenssista (SVST) minimaaliseen konvergenssiin, joka on samalla maksimaalinen divergenssi (EVET). EVET- ja SVET-sanaryhmien konvergenssiaste on sama, mutta konvergenssin ja divergenssin yhdistelmä on erityyppinen. Tämä antaa mahdollisuuden selvittää, onko taivutusmuotojen käytön kannalta semanttinen vai morfologinen divergenssi ratkaisevampi.

Aineiston analyysistä kävi ilmi, että sekä S2- että E2-oppijoiden teksteissä esiintyy vielä yksi, kieltenvälisen samanlaisuuden näkökulmasta mielenkiintoinen sanaryhmä (LVET): sanat, joilla on virossa ja suomessa

läheinen vartalo ja sama merkitys, mutta eri taivutusmalli, joka kuitenkin toimii mallina molemmissa lähdekielissä vaikka ei samoissa vaan eri sanoissa. Tämänäyttöisiä sanoja ovat esimerkiksi molemmissa kielissä samamerkityksiset suomen *elämästä* ja viron *elust* tai suomen *luonnosta* ja viron *loodusest*. Vartalot ovat fonologiselta ulkoasultaan sen verran läheisiä, että oppijat pystyvät havaitsemaan muotojen samanlaisuuden ja hyödyntävät oletettua samanlaisuutta myös tuottamisprosessissa. Vaikka taivutusmalli on näissä sanoissa molemmissa kielissä erilainen, on kohdekielen taivutusmalli oppijoille lähdekielen muiden sanojen pohjalta tuttu: esim. suomen *elämästä* taipuu kuten viron *sadamast* ja viron *elust* kuten suomen *lelusta*, suomen *luonnosta* kuten viron *linnust* ja viron *loodusest* kuten suomen *lautasesta*. Kuviosta 1 käy ilmi, että LVET-ryhmän elatiivimuotojen osuus on varsinkin E2-oppijoiden, mutta myös S2-oppijoiden korpusteksteissä melko merkittävä.



KUVIO 1. E2-oppijoiden ja S2-oppijoiden yksikön elatiivin muotojen jakauma oppijanvirron ja oppijansuomen korpuksissa viron ja suomen elatiivimuotojen samanlaisuuden ja erilaisuuden perusteella

Tämä tosiasia vahvistaa taas kerran aikaisempia huomioita (Ringbom 2007; Sajavaara 2006; Kaivapalu 2004; 2005) siitä, että havaitulla ja oletetulla samanlaisuudella on kielenoppijan prosessoinnissa todellisen, kieltenvälisen rakenteellisen samanlaisuuden rinnalla merkittävä rooli. Suurin on LVET-sanaryhmän osuus E2-oppijoiden B1-tason teksteissä, joissa esiintyykin vain virossa ja suomessa morfologisesti konvergoivien sanaryhmien sanoja. S2-oppijat ovat B1-tason teksteissä käyttäneet eniten taivutusmuotoja, joissa viron ja suomen konvergenssi on maksimissaan, siis semanttinen, morfologinen ja fonologinen. Jonkin verran esiintyy myös sanoja, joissa on virossa ja suomessa sama fonologinen rakenne ja taivutusmalli, vaikka niillä on eri merkitys. Sen sijaan virossa ja suomessa eri tavalla taipuvia elatiivimuotoja ei käytetä E2-oppijoiden B1-tason teksteissä ollenkaan, ei edes sellaisia, jotka ovat molemmissa kielissä merkitykseltään ja fonologiselta rakenteeltaan samanlaisia. Tuloksista ilmenee siis, että E2-oppijoiden elatiivimuotojen tuottamisprosessi B1-tason teksteissä pohjautuu selvästi lähde- ja kohdekielen taivutusmallien samanlaisuuteen eli morfologiseen konvergenssiin. Jos morfologiseen konvergenssiin yhdistyy semanttinen tai varsinkin semanttinen ja fonologinen konvergenssi, on kieltenvälinen samanlaisuus ilmeisesti oppijoille hyvin läpinäkyvä, helposti havaittavissa oleva ja toimii tuottamisprosessissa oletetun samanlaisuuden pohjana.

Vaikka E2-oppijoiden B2-tasonkin tekstien elatiivimuotojen käyttö perustuu valtaosin taivutusmallien samanlaisuuteen, voi silti huomata tuottamisprosessin laajenemista erilaisuuden suuntaan. B1-tason teksteihin verrattuna SVET- ja LVET-ryhmän sanojen käyttö on B2-tasolla hieman vähentynyt ja EVST-ryhmän sanojen käyttö puolestaan lisääntynyt. Uutena ryhmänä ovat tulleet mukaan EVET-ryhmän sanat, jossa viron ja suomen divergenssi on maksimissaan.

S2-oppijoiden B1- ja B2-tason teksteissäkin on eniten sanoja, jotka konvergoivat virossa ja suomessa fonologisesti, morfologisesti ja semanttisesti, eli SVST-ryhmän sanoja, vaikka näiden osuus on pienempi kuin E2-oppijoiden teksteissä. Toisin kuin E2-oppijoiden teksteissä on S2-oppijoiden teksteissä sekä B1- että B2-tasolla mukana kaikkien

sanaryhmien elatiivimuodot, myös ne, joissa on suomessa ja virossa eri taivutusmalli. E2-oppijoihin verrattuna on S2-oppijoiden teksteissä selvästi suurempi erivartaloisten samalla tavalla taipuvien sanojen eli EVST-ryhmän sanojen osuus. S2-oppijoille näyttää siis suomen ja viron taivutusmallien samanlaisuus olevan läpinäkyvämpi kuin E2-oppijoille. S2-oppijoiden elatiivinkäyttö laajentuu erilaisuuden suuntaan siis E2-oppijoihin verrattuna hieman eri tavalla: B2-tasolla kasvaa varsinkin LVET-ryhmän sanojen osuus.

Tämän tutkimuksen tulokset tukevat siis monelta osin aikaisempia vironkielisten suomenoppijoiden taivutustestin tuloksiin pohjautuvia tuloksia (Kaivapalu 2005). Ensinnäkin S2-oppijat hyödyntävät myös spontaanissa kirjallisessa kielenkäytössä lähde- ja kohdekielen taivutusmallien samanlaisuutta sekä fonologisen ja semanttisen samanlaisuuden tukemana että ilman sitä: noin kolme neljäsosaa kaikista oppijoiden tuottamista elatiivimuodoista perustuu nimenomaan taivutusmallien samanlaisuuteen. Toiseksi, erivartaloiset ja eri tavalla taipuvat sanat ovat oppijoille vähemmän haasteellisia kuin samavartaloiset eri tavalla taipuvat sanat, joiden fonologinen ja semanttinen samanlaisuus johtaa joko oppijan myös taivuttamaan sanoja lähdekielen mallin mukaan, kuten taivutustestin tulokset osoittivat, tai käyttämään näitä vähemmän kuin muiden sanaryhmien sanoja.

5. Pohdintoja, haasteita ja jatkumahdollisuuksia

Tutkimustulokset osoittavat, että valtaosa sekä E2-oppijoiden että S2-oppijoiden tuottamista elatiivimuodoista EVKK:n ja ICLFI:n teksteissä perustuu lähde- ja kohdekielen taivutusmallien samanlaisuuteen. Voidaan siis todeta, että morfologinen lähdekielen positiivinen vaikutus lähisukukielen nominien taivutusmuotojen omaksumiseen on symmetristä ja toimii sekä E2- että S2-oppijoiden kirjallisessa tuottamisprosessissa. Tämä heijastuu oppijoiden teksteissä käyttämien taivutusmuotojen valinnassa. Näin merkittävä valinnan keskittyminen lähde- ja kohdekielessä samanlaisiin taivutusmuotoihin saattaa olla osoitus siitä,

että näiden muotojen suhteellisen helpolla prosessoimilla säästetään työmuistin kapasiteettia muuhun tuottamistoimintaan.

Kieltenvälinen samanlaisuus ei kuitenkaan ole homogeeninen ilmiö vaan jatkumo maksimaalisesta konvergenssista maksimaaliseen divergenssiin, jossa on erityyppisiä konvergenssin ja divergenssin yhdistelmiä. Mitä enemmän leksikaalis-kieliopilliset rakenteet konvergoivat, sitä enemmän oppijat käyttävät niitä teksteissä. Sekä S2- oppijat että E2-oppijat hyödyntävät eniten fonologisen, morfologisen ja semanttisen samanlaisuuden yhdistelmää, mutta myös fonologisen ja morfologisen samanlaisuuden yhdistelmää. Merkittävää on myös sellaisten taivutusmuotojen käyttö, joissa on molemmissa kielissä läheinen vartalo ja erilainen, mutta sekä lähde- että kohdekielessä esiintyvä taivutusmalli. Ratkaiseva on siis taivutusmallien samanlaisuus, morfologinen konvergensi. Lähde- ja kohdekielessä samanlaisten muotojen rinnalle kehittyvät vähitellen erivartaloiset ja eri tavalla taipuvat muodot ja vasta sen jälkeen samanvartaloiset eri tavalla taipuvat muodot.

Lähisukukielen vaikutus viron ja suomen tuottamisprosesseissa ei kuitenkaan ole identtistä. S2-oppijoiden elatiivimuotojen jakauma on E2-oppijoiden jakaumaa heterogeenisempi: sekä B1- että B2-tason teksteissä esiintyy virossa ja suomessa samanlaisten muotojen lisäksi eri tavalla taipuvia taivutusmuotojakin, vaikka niiden osuus kokonaismäärästä ei ole suuri. S2-oppijoiden teksteissä on suomenkielisiin oppijoiden teksteihin verrattuna enemmän virossa ja suomessa fonologisesti ja morfologisesti konvergoivia, mutta semanttisesti divergoivia muotoja. Lähde- ja kohdekielen eri vartalon piilottama taivutusmallin samanlaisuus näyttää siis olevan S2-oppijoille läpinäkyvämpi kuin E2-oppijoille, jotka tarvitsevat semanttisen konvergenssin tuen havaitakseen taivutusmallin konvergenssia. Sen ilmiön syiden perusteellisempi selvittely tarvitsee tähän artikkeliin mahtuvaa tarkastelua laajemman analyysin.

S2- ja E2-oppijoiden B1- ja B2-tason tekstien elatiivimuotojen käytön vertailusta näkyy selvästi kielitaidon kehittyminen kieltenvälisestä samanlaisuudesta erilaisuuden suuntaan. S2- ja E2-oppijoiden tuotos-ten erojen selvittäminen vaatii jatkossa samojen kysymysten testaamista

isommalla aineistolla, jonka tekstejä on yleiseurooppalaisen viitekehyyksen taitotasojen kuvausasteikon mukaisesti arvioinut vähintään kolme arvioijaa. Jatkotutkimuksen suurimpia haasteita ovatkin viron oppijankielen korpuksen suomenkielisten oppijoiden aineiston laajentaminen ja molempien korpusten tekstien jatkuva arviointi. Korpuspohjaista lähdekielen vaikutuksen tutkimusta edistäisi myös korpusaineiston morfosyntaktinen koodaus, joka oppijansuomen korpuksen vironkielisten aineistossa alkaa olla käytettävissä. Tämän tutkimuksen tulokset osoittavat selvästi, että pelkkä virhekoodaus ei riitä, sillä kaikki pilottikorpuksen oppijoiden tuottamat elatiivimuodot ovat kohdekielen normin kannalta hyväksyttäviä.

Lähdekielen vaikutuksen tutkimisen yhdistetyn metodologisen viitekehyyksen (engl. *unified methodological framework*; Jarvis 2000) mukaan riittävä lähdekielen vaikutuksen todistusaineisto edellyttää jatkossa tulosten vertailua muiden lähdekielten käyttäjien tuloksiin sekä viron ja suomen natiiviaiaineistoon. Jatkuva haaste on myös todellisen samanlaisuuden psykolingvistisen perustan eli todellisen, havaitun ja oletetun samanlaisuuden suhteen selvittäminen: millaiset kielijärjestelmien samanlaisuudet ovat samanlaisuuksia oppijoidenkin mielestä ja millaiset eivät, millaiset samanlaisuudet ovat enemmän ja millaiset vähemmän samanlaisia, ovatko oppijoiden mielestä samanlaisia esimerkiksi viron *mehest* ja suomen *miehestä*, viron *muusikast* ja suomen *musiikista*, viron *asjust* ja suomen *asioista*.

Tämän pilottitutkimuksen tulokset näyttävät lupaavilta menetelmällisestikin: kahden lähisukukielen vaikutuksen symmetrinen tarkastelu toisaalta tukee yksisuuntaisten kieliparettain toteutettujen tutkimusten tuloksia, mutta toisaalta osoittaa myös, että kahden lähisukukielen vaikutus toistensa omaksumiseen ei ole välttämättä identtisesti symmetristä. Tämän ilmiön jatkotarkastelu saattaa laajemman ja metodologisesti tarkemman aineiston pohjalta valaista paljon tutkitun lähdekielen vaikutuksen luonnetta uudesta näkökulmasta.

Kiitokset

Tutkimusta ovat rahoittaneet Viron Tiedesäätiö (hankkeet ETF 8240 *Lähtekeele mõju ja teise keele omandamine: korpuspõhine uurimus* ja MJD104 *Symmetry of the cross-linguistic influence in the acquisition of closely related languages*), Viron valtakunnalliset ohjelmat *Viron kieli ja kulttuurimuisti (REKKi käsikirjaliste materjalide digiteerimine* ja *Eesti vahekeele korpuse alamkorpuste loomine ja korpuse kasutusvõimaluste populariseerimine*) ja *Viron kielen kieliteknologinen tuki (VAKO: Eesti vahekeele korpuse keeletarkvara ja keeletehnoloogilise ressursi arendamine)* sekä Ruotsin valtiopankin juhlarahasto.

Lähteet

- Bohnacker, Ute, Christina Rosén 2008. The clause-initial position in L2 German declaratives: Transfer of information structure. – *Studies in Second Language Acquisition* 30 (4), 511–538. doi:10.1017/S0272263108080741
- Eslon, Pille 2007. Õppijakeelekorpused ja keeleõpe. – Pille Eslon (Toim.). Tallinna Ülikooli keelekorpusete optimaalsus, töötlemine ja kasutamine. Tallinna Ülikooli eesti filoloogia osakonna toimetised 9. Tallinn: TLÜ Kirjastus, 87–120.
- Eslon, Pille, Helena Metslang 2007. Õppijakeel ja eesti vahekeele korpus. – *Eesti Rakenduslingvistika Ühingu aastaraamat* 3, 99–116.
- Eurooppalainen viitekehys 2003. Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys. Helsinki: WSOY.
- Granger, Sylviane 1998. The computer learner corpus: A versatile new source of data for SLA research. – Sylviane Granger (Ed.). *Learner English on Computer*. London: Longman, 3–18.
- Granger, Sylviane, Fanny Meunier 2010. Second language acquisition research and learner corpus research: Friend or foe? – Paper presented in 20th Annual Meeting of the European Second Language Association.
- Rasier, Laurent, Philippe Hilgsmann 2007. Prosodic transfer from L1 to L2. Theoretical and methodological issues. – *Nouveaux cahiers de linguistique française* 28, 41–66.
- Jantunen, Jarmo H. 2007. Oppijansuomen piirteitä korpusvetoisesti. – Pirkko Muikku-Werner, Ossi Kokko, Hannu Remes (Toim.). *Virsu III. Suomalais-ugrilaisia kohdekieliä ja kontakteja*. *Studies in Languages* 42. Joensuu: Joensuun yliopisto, 69–83.

- Jantunen, Jarmo H. 2008. Haasteita oppijankielen korpusanalyyseille: oppijankielen universaalit. – Pille Eslon (Toim.). Öppijakeele analüüs: võimalused, probleemid, vajadused. Tallinna Ülikooli eesti filoloogia osakonna toimetised 10. Tallinn: Tallinna Ülikool, 67–92.
- Jantunen, Jarmo H. 2009. *Minulla on aivan paljon rahaa* – Fraseologiset yksiköt suomen kielen opetuksessa. – Virittäjä 3, 356–381.
- Jantunen, Jarmo H. 2010. Avainsana-analyysi annotoidun oppijankieliaineiston tutkimuksessa: alustavia havaintoja. – Esitelmä AFinLA:n syysseminariumin työpajassa Metodit L2-korpustutkimuksessa.
- Jantunen, Jarmo H., Saana Piltonen 2009. Oppijansuomen ja -viron sähköiset tutkimusaineistot. – Virittäjä 113, 449–458.
- Jarvis, Scott 2010. Comparison-based and detection-based approaches to transfer research. – Leah Roberts, Martin Howard, Muiris Ó Laoire, David Singleton (Eds.). EUROSLA Yearbook 10. Amsterdam: Benjamins, 169–192. [doi:10.1075/eurosla.10.10jar](https://doi.org/10.1075/eurosla.10.10jar)
- Jarvis, Scott, Aneta Pavlenko 2008. Crosslinguistic Influence in Language and Cognition. New York, London: Routledge.
- Jeong, Hyeonjeong, Motoaki Sugiura, Yuko Sassa, Satoru Yokoyama, Kaoru Horie, Shigeru Sato 2007. Crosslinguistic influence on brain activation during second language processing: An fMRI study. – Bilingualism: Language and Cognition 10 (2), 175–187. [doi:10.1017/S1366728907002921](https://doi.org/10.1017/S1366728907002921)
- Kaivapalu, Annekatrin 2004. Kui sarnane on sarnane? Eesti ja soome mitmusevormide psühholingvistilisest reaalsusest. – Helena Sulkala, Heli Laanekask (Toim.). VIRSU II. Suomi ja viro kohdekielinä. Oulun yliopiston suomen ja saamen kielen ja logopedian laitoksen julkaisu 24. Oulu: Oulun yliopisto, 62–71.
- Kaivapalu, Annekatrin 2005. Lähdekieli kielenoppimisen apuna. Jyväskylä Studies in Humanities 44. Jyväskylä: Jyväskylän yliopisto.
- Kaivapalu, Annekatrin 2008. Lähtekeele mõju korpuspõhine uurimine. – Pille Eslon (Toim.). Öppijakeele analüüs: võimalused, probleemid, vajadused. Tallinna Ülikooli eesti filoloogia osakonna toimetised 10. Tallinn: Tallinna Ülikool, 93–119.
- Kaivapalu, Annekatrin 2009. Vironkielisen suomenoppijan äidinkieli – ongelma, haaste vai voimavara? – Virittäjä 3, 382–401.
- Kallioranta, Otto 2009. Paljon-adverbin kollogointi oppijansuomessa. Korpusvetoinen tutkimus. Julkaisematon pro gradu -tutkielma. Oulun yliopisto. <http://www.oulu.fi/hutk/sutvi/oppijankieli/tutkimus> (17.2.2011).

- Kotz, Sonja A. 2009. A critical review of ERP and fMRI evidence on L2 syntactic processing. – *Brain & Language* 109 (2–3), 68–74. [doi:10.1016/j.bandl.2008.06.002](https://doi.org/10.1016/j.bandl.2008.06.002)
- Nesselhauf, Nadja 2003. The use of collocations by advanced learners of English and some implications for teaching. – *Applied Linguistics* 24 (2), 223–242. [doi:10.1093/applin/24.2.223](https://doi.org/10.1093/applin/24.2.223)
- Nesselhauf, Nadja 2005. *Collocations in Learner Corpus*. Amsterdam: John Benjamins.
- Odlin, Terrence 1989. *Language Transfer. Cross-linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Ringbom, Håkan 2007. *Cross-linguistic Similarity in Foreign Language Learning*. Clevedon: Multilingual Matters LTD.
- Sajavaara, Kari 2006. Kontrastiivinen analyysi, transfer ja toisen kielen oppiminen. – Annetkatrin Kaivapalu, Külvi Pruuli (Toim.). *Lähivertailuja* 17. Jyväskylä Studies in Humanities. Jyväskylä: Jyväskylän yliopisto, 9–26.
- Tognini-Bonelli, Elena 2001. *Corpus Linguistics at Work*. Amsterdam: Benjamins.
- Tokowicz, Natasha, Brian MacWhinney 2005. Implicit and explicit measures of sensitivity to violations in second language grammar. – *Studies in Second Language Acquisition* 27 (2), 173–204. [doi:10.1017/S0272263105050102](https://doi.org/10.1017/S0272263105050102)
- Trofimowich, Pavel, Elizabeth Gatbonton, Norman Segalowitz 2008. A dynamic look at L2 phonological learning: Seeking processing explanations for implicational phenomena. – *Studies in Second Language Acquisition* 29 (3), 407–448. [doi:10.1017/S027226310707026X](https://doi.org/10.1017/S027226310707026X)

Annetkatrin Kaivapalu

Jyväskylän yliopisto, soveltavan kielitutkimuksen keskus
PL 35
40014 Jyväskylän yliopisto, Finland

Tallinnan yliopisto, viron kielen ja kulttuurin instituutti
Narva mnt 25
10120 Tallinn, Estonia
kaivapa@tlu.ee

Pille Eslon

Tallinnan yliopisto, viron kielen ja kulttuurin instituutti
Narva mnt 25
10120 Tallinn, Estonia
peslon@tlu.ee

Is the influence of a closely related first language on the acquisition of Finnish and Estonian symmetrical? Some preliminary results and challenges of a corpus-based study

ANNEKATRIN KAIVAPALU

Jyväskylä University, Tallinn University

PILLE ESLON

Tallinn University

The paper deals with the influence of closely related first language on the acquisition of Finnish and Estonian. This study is a part of the research project *Cross-linguistic influence and second language acquisition: corpus-based research of Tallinn University* and is based on parallel use of two learner language corpora, the Estonian Interlanguage Corpus (EILC) in Tallinn University (<http://evkk.tlu.ee>) and the International Corpus of Learner Finnish (ICLFI) in Oulu University (<http://www.oppijankieli.fi>). One of the main research questions of the project addresses relationships between objective similarity and symmetry of the cross-linguistic influence: is positive cross-linguistic influence also symmetrical, when symmetrical objective similarity occurs between the systems of L1 and L2?

This paper focuses on symmetry of positive morphological first language influence on the bases of the examples of EILC and ICLFI which were chosen according to convergence and divergence between Estonian and Finnish morphological structures. Five groups of elative singular forms in texts of proficiency levels B1 and B2 according to a scale of *Common European Framework of Reference for Languages: Learning, teaching and assessment* (CEFR), written by Estonian learners of Finnish and Finnish learners of Estonian were analyzed:

- 1) words with the same or a very similar stem, meaning and inflection in Finnish and Estonian;
- 2) words with a different stem and meaning, but similar inflectional pattern;

- 3) words with close stem, the same meaning, but different inflectional pattern which exists in other words of the learners' L1;
- 4) words with the same or a very similar stem and meaning in Finnish and Estonian, but with different inflection;
- 5) words with both a different stem and inflection.

The results of the study suggest that positive morphological influence of a closely related first language on the producing process of Finnish and Estonian is symmetrical and acts in both directions. Both Estonian learners of Finnish and Finnish learners of Estonian clearly prefer to use inflectional forms of maximum similarity, i.e. of phonological, morphological and semantical convergence in L1 and L2. The morphological convergence, similarity of inflectional patterns, turned out to be more critical a factor in terms of morphological cross-linguistic influence than semantical convergence, the similarity of word meaning. The similarity of inflectional patterns is more salient for Estonian learners of Finnish than for Finnish learners of Estonian, whose perception of inflectional pattern similarity needs to be supported by phonological and semantical similarity. The production of elative singular forms by Finnish learners of Estonian shows clear development from cross-linguistic similarity to differences, while the production of Estonian learners of Finnish is more heterogeneous and includes all five groups of elative singular forms. However, the percentage of elative forms with different inflection patterns in Finnish and Estonian is marginal.

The paper also discusses some challenges of corpus-based research of cross-linguistic influence such as reliability of the data and of proficiency level evaluation, **synchronizing two learner language corpora** and the psycholinguistic reality of cross-linguistic similarity.

Keywords: second language acquisition; corpus-based research; cross-linguistic influence; closely related languages; convergence; divergence; morphology; nominal inflection; Finnish; Estonian