

# Mitä pitkittäistutkimus paljastaa edistyneiden suomenoppijoiden kielitaidosta?

KIRSTI SIITONEN, JENNY NIEMELÄ

Turun yliopisto

**Tiivistelmä.** Tässä artikkelissa tarkastellaan kolmen edistyneen suomenoppijan kielitaitoa suomen ja sen sukukielten maisteriopintojen alussa ja 7 kuukautta opintojen alkamisesta. Artikkelissa esitellään, miten tietyt indikaattorit käyttäytyvät tutkimusajan kohtina, ja pohditaan, mitä ne voivat paljastaa kielitaidosta ja sen muutoksista.

Pitkälle edistyneen kielenoppijan kielitaito näyttyy hyvin vaihtelevana eri tarkasteluajankohtina. Eroihin voivat kuitenkin olla vaikuttamassa tekstin luonteesta tai sisällön leksikaalisista vaatimuksista johtuvat seikat. Tällöin kyse saattaaakin olla samankaltaisista ilmiöistä kuin äidinkielisillä puhujilla eikä suinkaan kielitaidon rajallisuudesta.

Todennäköistä ei ole, että eri oppijoiden kielitaito olisi samanlainen vastaavina tarkasteluajankohtina. Tutkimuksessa olisikin arvokkainta onnistua havaitsemaan oppijoille yhteiset muutostendenssit ja ne tyypilliset erot, joita on edistyneidenkin akateemisten oppijoiden ja äidinkielisten puhujien välillä. Kvantitatiivisen analyysin perusteella predikaatteihin liittyvät ongelmat vähenevät informanteilla. Muotovirheet eivät näytä kuitenkaan vähenevän kielitaidon kohentuessa, eikä infiniittisiä rakenteita ole käytössä niin runsaasti kuin äidinkielisillä puhujilla. Holistisessa kvalitatiivisessa analyysissä todetaan, että jälkimmäisenä tarkasteluajankohtana kaikkien informanttien tekstit ovat erittäin ymmärrettävää kieltä ja korrekkeja kokonaisuuksia esiintyy jokaisella.

**Avainsanat:** akateeminen oppija; kielenkehityksen indikaattorit; kirjoitusvirheet; korpustutkimus; muutostendenssi; tenttivastaukset; suomi toisena kielenä; verbien hallinta

## 1. Johdanto

Kaikki suomalais-ugrilaiset kielet, ehkä unkaria lukuun ottamatta, ovat melko pienten kansanryhmien puhumia. Tämän vuoksi äidinkielliset puhujat eivät välttämättä ole tottuneet siihen, että joku osaa heidän kieltään erinomaisesti akateemisella tasolla, eivätkä he myöskään osaa tätä vaatia. Se, että ymmärrettäisiin suomen kielen akateemisen kielitaidon arvo ja merkitys, osoittaisi, että opiskelevaa yksilöä kunnioitetaan, omaa äidinkieltä arvostetaan ja asian yhteiskunnallinen merkitys Suomessa tajutaan. Erinomaista akateemista suomea tarvitaan myös Suomen ulkopuolella, kun esimerkiksi yliopistonopettaja, joka ei ole syntyperäinen suomen puhuja, opettaa suomea korkealla tasolla. Pitkällä olevien suomenoppijoiden kieltä on siis aiheellista tutkia.

Tässä artikkelissa selvitetään edistyneen, akateemisen suomenoppijan kieltä tarkastelevan pitkittäistutkimuksen mahdollisuuksia ja keinoja sekä myös sitä, mitä tutkimuksella voidaan saada selville. Perusedellytyksenä on luonnollisesti mahdollisuus seurata pidemmän ajan kuluessa yhtä henkilöä tai yhtä ryhmää. Välttämättä on saatava myös riittävästi materiaalia.

Pitkittäistutkimuksessa saadaan selville yksilön todellinen reaaliaikainen kehitys; näin ei jouduta turvautumaan vain tiettyjen kielitaitotasojen erillisiin yksilö- tai yleiskuvauksiin (Ellis & Barkhuizen 2005: 97). Vastaavasti variaationtutkimuksessa voidaan käyttää reaaliaikametodia erotukseksi näennäisaikametodista (ks. esim. Kurki 2005: 36–37). Oppijankielessä nähdään mahdollinen variaatio ja sen muutos sekä koko kielitaidon muutos. Välikielen ilmiöthän ovat koko ajan liikkeessä.

Omassa tutkimuksessamme on kyse Turun yliopiston entisen suomalaisen ja yleisen kielitieteen – nykyisen kieli- ja käänntieteen – laitoksen suomen ja sen sukukielten maisteriohjelman opiskelijoista,

joiden suomen kielen taidon kirjallista kehitystä on mahdollista seurata vähintään kahden vuoden ajan. Nyt kyseessä olevassa pitkittäistutkimuksessa seurannan alkaessa tutkittavien kielitaito on melko hyvä. Tämän vuoksi hyvin nopeaa kehitystä ei ehkä ole odotettavissa. Ei liene myöskään todennäköistä, että kielitaito olisi loppuvaiheessa huomattavasti alkuvaiheen kielitaitoa parempi, mutta todennäköisesti erilainen.

Turun yliopistossa ollaan luomassa edistyneiden suomenoppijoiden koodattua korpusta, jonka avulla tulee olemaan mahdollista nähdä laajempien materiaalien variaatiota ja variaationkehitystä.

Korpuksen avulla voidaan tutkia myös äidinkielisten puhujien vastaavia tekstejä, joita korpuksessa on vertailuaineistona. (*Edistyneiden suomenoppijoiden korpus* -hanke LAS2, ks. esim. Ivaska & Siitonen 2009; Jantunen & Piltonen 2009.)

Tämä artikkeli toimii pilottina testaamassa korpuksemme käyttökelpoisuutta pitkittäistutkimukseen (ks. Ellis 1994: 75–77). Olemme ottaneet mitattavaksi eräitä muuttujia korpusvetoisesti. Korpusvetoisuudella tarkoitamme tässä, että korpusta luotaessa olemme huomanneet vain lukemalla, emme siis mittaamalla, tiettyjen ilmiöiden toistuvan tai olevan yllättävän frekventtejä oppijoiden korpusaineistossa. Tässä pilotissa tarkistamme, onko huomiollamme todellisuus pohjaa. Opetuskokemuksen pohjalta taas olemme valinneet toiset indikaattorit, joita tässä tutkimme siis korpuspohjaisesti. (Ks. esim. Tognini-Bonelli 2001: 84–85.)

## 2. Tutkittavana edistyneen oppijan kielitaidon kehitys

Edistyneiden oppijoiden teksteistä voidaan tutkia lähes natiivin puhujan tasolla olevan suomenoppijan kielitaidon luonnetta (vrt. Ringbom 1993). Lisäksi voidaan seurata, kuinka kielitaito kehittyy vapaan variaation ja uudelleenjärjestelyn kautta kohdekielenkaltaiseen taitoon (Ellis 1985: 95–96).

Tässä artikkelissa tarkastelemme kolmen edistyneen suomenoppijan kielitaitoa suomen ja sen sukukielten maisteriopintojen alussa ja

7 kuukautta opintojen alkamisesta. Artikkelissa esittelemme, miten tietyt indikaattorit käyttäytyvät tutkimusajankohtina ja miten ne liittyvät toisiinsa, ja pohdimme, mitä ne voivat paljastaa kielitaidosta. Yhdeltä informantilta on lisäksi aineistoa vaiheesta, jolloin hän on opiskellut ohjelmassa intensiivisesti jo 14 kuukautta. Tämä lisäaineisto tuo artikkeliin tiettyä epäyhtenäisyyttä, mutta nähdäksemme se toisaalta antaa arvokasta lisätietoa oppimisen etenemisen epätasaisuudesta. Eri informanteilta esittelemme myös osin eri ilmiöitä kunkin informantin tekstien erityislaadun perusteella.

Tutkimuksessamme etsimme mahdollista muutoksen suuntaa. Tutkimuksemme pyrkii löytämään siis kielenpiirteitä, jotka voisivat toimia indikaattoreina edistyneiden oppijoiden kielenkehityksestä. Myös mahdollinen muutosten puuttuminen kertoo oppijan kielen luonteesta. Tutkimuksemme aineisto on melko pieni, ja niinpä tutkimustamme voisikin kuvata tapaustutkimukseksi, jonka tulokset kertovat yksittäisten kirjoittajien eri teksteistä ja antavat viitteitä siitä, miten tekstit poikkeavat äidinkielisten teksteistä ja mitä piirteitä ja ilmiöitä kannattaisi tutkia isommista aineistoista, esimerkiksi korpustutkimuksen menetelmin. (Vrt. Ivaska & Siitonen 2010.)

Tutkimme ensin kvantitatiivisesti tiettyjä ilmiöitä ja sen jälkeen kvalitatiivisen lähiluvun avulla etsimme perusteluita ja taustaa kvantitatiivisiin tuloksiin. Luvuissa 3.3 ja 3.5 pyrimme näyttämään, mitä kvantitatiivinen analyysi pystyy osoittamaan kolmen tutkitun informantin kahden eri ajankohdan tekstien eroista. Sen jälkeen luvussa 3.6 on lyhyt kvalitatiivinen analyysi informanttien käyttämistä verbeistä ja erityisesti kunkin informantin teksteistä, jotta lukija saa paremman käsityksen siitä konkreettisesta, kokonaisista lauseista sisältävästä kielestä, jota informantit kirjoittavat.

Tutkimuskysymyksiksi ovat muotoutuneet seuraavat:

- 1) Onko havaittava muutos mahdollinen myös edistyneiden oppijoiden tasolla, ja kuinka rajallista kielitaidon kehittyminen tässä vaiheessa on?

- 2) Ovatko muutostendenssit nähtävissä havainnollisesti ja verrattavissa keskenään, jos ja kun lähtötilanne vaihtelee eri informanttien välillä?
- 3) Voidaanko muutosaikaa käyttää yhdistävänä tekijänä lähtötilanteesta riippumatta?

Näistä kysymyksistä nousee esiin myös laajempaa korpusta koskeva kysymys:

- 4) Koska kultakin informantilta on laajemmassakin korpuksessa vain rajallinen määrä aineistoa, yhdeksi tutkimuskysymykseksi nousee, miten sitoa eri informanttien aineistoja toisiinsa määrällisesti merkittävien otantojen aikaansaamiseksi.

### **3. Analyysi kolmen edistyneen oppijan tenttivastauksista**

#### **3.1. Aineisto, tutkittavat ilmiöt ja tutkimuksen hypoteesit**

Artikkelin aineistona on informanttien tenttivastauksia. Tenttivastausten käyttöä tutkimuksessa on toisaalta pidetty järkevänä, toisaalta kritisoitu. Tenttivastaukset ovat siinä mielessä spontaania ja autenttista aineistoa, että niitä kirjoittaessaan informantit eivät todennäköisesti keskity kieleen vaan nimenomaan asiasisältöön. Toisaalta eri henkilöillä on erilainen stressinsietokyky eikä tentissä todennäköisesti ole aikaa kielenkorjailuun. Näin tenteissä saattaa esiintyä virheitä, jotka kirjoittaja pysyisi korjaamaan rauhallisemmassa tilanteessa. Tällaisessa tutkimuksessa ei kuitenkaan pyritä saamaan selville, mitä kirjoittaja tietää kieliopista ja korrekteista muodoista vaan miten hän niitä spontaanisti käyttää. (Ks. myös Ellis–Barkhuizen 2005: 28–30.) Lisäpohdintoja tenttivastausten soveltuvuudesta tutkimusaineistoksi on luvussa 4.1.

Tenttivastaukset on alun perin kirjoitettu käsin. Olemme esittäneet tutkimuksen kuluessa muutaman huomion myös siitä, miten se, että tenttivastaukset on kirjoitettu käsin voi vaikuttaa siihen käsitykseen, jonka lukija saa tekstistä. (Vrt. Tarnanen 2003: 240.) Erityisesti tällai-

sessä tutkimuksessa, jossa ei olla antamassa arvosanaa kieliopillisesta oikeellisuudesta vaan tutkitaan tekstin kokonaisuutta, myös käsialaan liittyvät ilmiöt voidaan ottaa huomioon. On kyse tekstin ja lukijan vuorovaikutuksesta aivan samalla tavalla kuin kasvokkaisessa keskustelussa, jossa yleensä puheen intonaatio ja kontaktin luontevuus ovat ensisijaisia suhteessa puhutun kielen muotojen korrektiuteen. Myös tenttivastauksessa, josta opettaja haluaa saada tiedon vain siitä, hallitseeko tenttijä tenttittävän oppimäärän, asian selväpiirteinen esittäminen sekä käsialan ja muotoilun siisteys ovat ensiarvoisen tärkeitä.

Tutkittavat ilmiöt olemme valinneet osin korpusvetoisesti, osin opettajan- ja tutkijankokemuksen perusteella korpuspohjaisesti (vrt. edellä). Korpusta digitoitaessa oli huomattu, että suuri osa juoksevan tekstin predikaattiverbeistä näytti olevan *olla*-verbin eri muotoja. Päätimme tutkia, onko näitä selvästi enemmän kuin äidinkielisillä kirjoittajilla.

Tunnetusti kielenoppijoilla, joskin myös äidinkielisillä kirjoittajilla, on kirjoituksissaan monenlaisia epäkorrekteja ilmauksia. Päätimme keskittyä tarkastelemaan vain seuraavanlaisia virheitä: muoto- ja kirjoitusvirheitä sekä predikaattien käyttöön liittyviä ongelmia.

Lisäksi nimenomaan akateemisessa suomessa pidämme kiinnostavana erilaisten infinitiivi- ja partisiippirakenteiden hallintaa, koska nämä opetetaan suomen kieltä opiskeleville tavallisesti kieliopintojen lopuksi. Oletettavaa on, että näiden monipuolisempi käyttö lisääntyy edistyneiden oppijoiden kielessä.

Varmaa ei ole, ovatko näin valitut piirteet ja ilmiöt juuri sellaisia indikaattoreita, jotka paljastaisivat jotakin olennaista kielitaidon kehityksestä ja muutoksesta. Joka tapauksessa edelleen osin korpusvetoisesti ja osin opettajan- ja tutkijankokemuksen perusteella korpuspohjaisesti rohkenemme esittää joitakin hypoteeseja mahdollisista muutoksista. Pitkittäistutkimuksemme tulee mahdollisesti paljastamaan maisteriopiskelijan kehityksestä 7 kk:n ajanjakson aikana seuraavaa:

- 1) Alussa 80 % juoksevan tekstin predikaattiverbeistä on *olla*-verbin eri muotoja, lopussa enää n. 50 %. Verbivaraston käyttö monipuolistuu.

- 2) Alussa kirjoitelmissa on muoto- ja kirjoitusvirheitä vielä noin 5 / 300 sanaa, lopussa enää noin 1 / 300 sanaa.
- 3) Alussa kirjoitelmissa on n. 20 %:ssa predikaateista joko sananvalintaan tai käyttöön liittyviä ongelmia, ajanjakson lopussa enää n. 10 %:ssa.
- 4) Alussa kirjoitelmissa on monipuolisia infinitiivi- ja partisiippi-rakenteita noin 1 / 300 sanaa, lopussa niitä on jo noin 5 / 300 sanaa.

### 3.2. Tutkimuksen informantit

**Informantti 1** (In1) on islantilainen nainen, joka on opiskellut suomea ainoastaan kahden vuoden ajan ennen muuttoa Suomeen. Informantilla 1 ei vielä tutkimuksen alkaessa ollut opiskelun ulkopuolisia suomenkielisiä kontakteja.

**Informantti 2** (In2) on venäläinen nainen, joka on opiskellut suomea Venäjällä neljä vuotta. Informantti 2 on informantin 1 tavoin muuttanut Suomeen vasta tutkimuksen alussa. Hänellä on kuitenkin suomenkielinen asuintoveri. Lisäksi hän opiskelee viroa, mikä oletettavasti tukee hänen suomen kielen opintojaan, sillä esimerkiksi suomen ja viron kieliopillisten ja sanastollisten yhtäläisyyksien ja eroavaisuuksien tarkasteleminen voi helpottaa joidenkin ilmiöiden ymmärtämistä ja omaksumista.

**Informantti 3** (In3) on japanilainen nainen, joka on opiskellut suomea ennen tutkimuksen alkua kaksi vuotta Japanissa ja viisi vuotta Suomessa. Informantilla on suomalainen puoliso ja suhteellisen paljon suomenkielisiä kontakteja. Itsearviointilomakkeessa hän arvelee kuitenkin hallitsevansa suomen kirjoittamisen paremmin kuin puhumisen.

**Verrokkeina** (Verrokki 1, Verrokki 2 ja Verrokki 3) on kolme äidinkielistä suomen kielen opiskelijaa. Vertailuaineistona on yksi tenttivastaus kultakin verrokilta. Näistä on etsitty samat seikat kuin informanttienkin teksteistä.

### 3.3. Analyysi tutkittavista ilmiöistä informanteittain

Kolmen informantin tenttivastauksista olemme tehneet<sup>1</sup> kvantitatiivisen analyysin tekstien saneiden ja lekseemien määristä sekä virkkeiden keskipituuksista. Sen jälkeen olemme keränneet verbeistä *olla*-verbin eri muodot ja samalla tarkastelleet verbiesiintymien monipuolisuutta sekä predikaattiin liittyviä ongelmia. Tämän lisäksi olemme poimineet kirjoitelmien kompleksiset infinitiivi- ja partisiippi rakenteet sekä muoto- ja kirjoitusvirheet. Kaikilta informanteilta on aineistoa kahdesta vaiheesta (A ja B); lisäksi informantilta 3 aineistoa on myös kolmannesta vaiheesta (vaihe C).

#### 3.3.1. Informantin 1 A- ja B-vaihe

Informantin 1 A-vaiheen tenttivastaukset koostuvat kahdesta tentistä, joissa on yhteensä 413 sanetta. Vertailukohtana on informantin seitsemän kuukauden kuluttua kirjoittama 694 saneen mittainen tenttivastaus. Informantin 1 A-vaiheen tenteissä lekseemien suhteellinen määrä on 149/300. Vaikka informantin 1 B-vaiheen tenttivastaus onkin huomattavasti A-vaiheen vastauksia pidempi ja virkkeiden keskipituus on noussut 13:sta 16:een, lekseemien suhteellinen määrä on toisen vaiheen tenttivastauksissa hieman alhaisempi (141/300). Toisaalta on otettava huomioon, että A-vaiheessa käsitellään kahta tenttiä, joiden eri aihepiirit jo saattavat tuoda tekstiin lisää eri lekseemejä. Muoto-opillisten virheiden suhteellinen määrä on pysynyt muuttumattomana, niitä on molempien vaiheiden tenteissä n. 7 kappaletta 300:aa sanaa kohti (A-vaiheessa 9 kpl ja B-vaiheessa 17 kpl). Joukossa on muun muassa useita astevaihteluun liittyviä virheitä, joita muilla informanteilla ei juuri esiinny (*etteivät vaati, taidonsa, muodojen, matella, valeta* [‘valjeta’], *ei sopi*).

Verbeissä informantin 1 kehitys näkyy kuitenkin selvemmin. *Olla*-verbin (muiden verbien liittomuotojen *olla*-verbejä ei ole laskettu mukaan) osuus predikaateista on laskenut peräti 18 prosenttiyksikköä:

<sup>1</sup> LAS2-korpuksen harjoittelija Sara Suominen on osallistunut poimintoihin.



A-vaiheen tenteissä predikaateista 52,5 % (31 kpl) on *olla*-verbin eri muotoja, toisen vaiheen tentissä enää 34,5 % (30 kpl). A-vaiheen tentti-vastauksissa on yhteensä 20 erilaista predikaattia ja B-vaiheen tentissä 31. A-vaiheen tenteissä suurin osa (7 esiintymää, 78 %) predikaattiverbien käyttöön liittyvistä ongelmista koskee rektiota. Esimerkissä (1) on kyse partitiiviverbiin, esimerkissä (2) taas fraasiin “puhua jotakin kieltä” liittyvästä ongelmasta.

- (1) Unkarilainen lääkäri ja kielitiedemies Samuel Gyármathi (1751-1830) käytti esimerkiksi Sajnovicsin **kirjoitukset** ja perusti siihen oma kirjoitus, fennoungriistien kielten vertailevan historian esittely (julk. 1799). (In1A)
- (2) **Se** puhutaan luoteis-Venäjällä. (In1A)

Lisäksi on sekamuodoste:

- (3) Minun tarvitsee sinut. (In1A)

B-vaiheen predikaatteihin liittyvät ongelmat eivät ole painottuneet millenkään tietylle alueelle – rektiovirheitäkin on vain kaksi. Virheellisten predikaattien prosentuaalinen osuus on laskenut puoleen siitä, mitä se A-vaiheen tenteissä oli (15 %:sta (abs. 9 kpl) 8 %:iin (abs. 7 kpl)). Seitsemän prosenttiyksikön ero on suhteellisen suuri, kun otetaan huomioon, että jäljempänä informantilla 3 ei kaksi kertaa pidemmällä aikavälillä tapahtunut kuin kahden prosenttiyksikön lasku. On kuitenkin syytä ottaa huomioon, että informanttien taso ja kehitysnopeus vaihtelevat tutkittavan kielenpiirteen mukaan: informantin 1 teksteissä on esimerkiksi useita subjektin ja objektin sijavalintaan liittyviä ongelmia (esim. *Eteläsamojedikielet on kaksi, Joku voi avata ovea*), joita informantin 3 teksteissä ei esiinny. Muu kuin rektioksi määriteltävissä oleva sijavalinta ei kuitenkaan kuulu tähän tutkimukseen valittuihin piirteisiin.

Infinitiivi- ja partisiippirakenteiden suhteellinen määrä on kasvanut (A-vaiheessa 8/300 (abs. 11 kpl), B-vaiheessa 11/300 (abs. 26 kpl)). Informantti 1 käyttää enemmän infinitiivi- ja partisiippirakenteita kuin esimerkiksi informantti 3, mutta rakenteet eivät ole yhtä monipuolisia. A-vaiheessa informantin tekstissä esiintyy ainoastaan A-infinitiivejä, VA-

partisiippeja ja yksi agenttipartisiippi (ks. asetelma 1). B-vaiheessa inventaari on monipuolisempi, vaikkakin A-infinitiivin osuus on peräti 68 % (19 kpl). Lisäksi tenttivastauksessa on 3 VA-partisiippia, 1 E-infinitiivi, 1 NUT-partisiippi, 1 MA-infinitiivi ja 1 TU-partisiippi (ks. asetelma 2).

A-infinitiivin sisältävistä rakenteista 15 edustaa rakennetta *voida* + A-inf. (esim. *voi selittää, voi tehdä, voi olla, eivät voi esiintyä, voivatkin muuttua*). Informantti 1 näyttääkin tukeutuvan ensisijaisesti yksinkertaisiin rakenteisiin. Apuverbiä ja A-infinitiiviä käyttämällä informantti 1 saattaa myös pyrkiä välttämään sellaisten verbien taivutusta, joiden taivutussäännöistä hän ei ole täysin varma. Päätelmää tukee se, että informantin 1 muoto-opilliset virheet ovat nimenomaan astevaihteluvirheitä. Toisaalta yksinkertaisenkin muodon valitseminen saattaa johtaa esimerkiksi rektiovirheeseen, kuten virkkeessä 4; tosin myös suomenkieliset saattavat puhemielessään ilmaista asian näin.

- (4) Proseduraalinen eli implisiittinen tieto on toisaalta se, kun oppija voi lausua oikeita ilmauksia kohdekielellä, vaikka ei hän välttämättä **psytyisi selittää** tietoaan tai kertoa siitä säännöllisesti. (In1B)

Uusien infinitiivi- ja partisiippirakenteiden käyttö on vielä jokseenkin epätarkkaa:

- (5) **Käytetään** samaa persoonapronominia **puhumalla** miehestä ja naisesta. (In1B)
- (6) Viimeiseksi **neljäs mainittu** suomen kielen ominaispiirteistä on äännevaihtelu. (In1B)

Myös VA-partisiipin käytössä on A-vaiheen tenttivastauksista poiketen ongelmia.

- (7) Ensimmäisen ryhmän vokaalit voivat esiintyä **molempien toisien sisältävien vokaalien** kanssa. (In1B)
- (8) Tästä voi abstrahoida, että kaks.tavuiset, e-loppuiset substantiivit monikon elatiivissa, ja **niiden johtavien -nen loppuiset adjektiivit** partitiivissa, ovat säännöllisesti homonyymit. ('niistä johdetut -nen-loppuiset adjektiivit'). (In1B)

**ASETELMA 1.** *Informantin 1 A-vaiheen infinitiivi- ja partisiippirakenteita*

<b>A-infinitiivi</b>	Sajnovics <b>halusi osoittaa</b> unkarin ja saamen kielten välisten suhteet vääräksi, sen takia että hän piti unkarin kieli saamen niin hyvin korkeampana.
<b>VA-partisiippi</b>	Ykkösargumentti puuttuu lauseilta 1 ja 4, joissa <b>näkyvää tekijää</b> ei ole – joku tuo koiran, joku voi avata ovea, mutta ei se näytä lauseilta.
<b>agentti-partisiippi</b>	Vaikka tämä hypoteesi oli tietysti väärin, <b>Sajnovicsin tekemää</b> tutkimusta tuli tärkeäksi seuraaville kielitutkijoille.

**ASETELMA 2.** *Informantin 1 B-vaiheen infinitiivi- ja partisiippirakenteita*

<b>A-infinitiivi</b>	Konsonantit <b>voivatkin muuttua</b> laadullisesti substantiivien taivutuksessa
<b>E-infinitiivi</b>	Puhutaan motivaation lajeista – välineellisestä motivaatiosta (e. instrumental motivation), kun oppijalla on <b>opiskellessaan</b> mielessään esim. tenttisuoritus tai työpaikka, tms. joka helpommin saa hyvällä kielentaidolla.
<b>MA-infinitiivi</b>	Toiseksi, suomesta puuttuu kieliopillinen suku – <b>käytetään</b> samaa persoonapronominia <b>puhumalla</b> miehestä ja naisesta.
<b>VA-partisiippi</b>	Ensimmäisen ryhmän vokaalit voivat esiintyä <b>molempien toisien sisältävien vokaalien</b> kanssa, mutta toisen ja kolmanteen ryhmien vokaalit eivät voi esiintyä yhdessä samassa sanassa (yhdyssanat tietenkin lukuun ottamatta).
<b>NUT-partisiippi</b>	Ideaalisesti, <b>Suomessa syntyneiden</b> ja sinne muuttavien määrä olisi tasaisempaa.
<b>TU-partisiippi</b>	Viimeiseksi <b>neljäs mainittu</b> suomen kielen ominaispiirteistä on äännevaihtelu.

Virke (9) on esimerkki A-vaiheen korrektista VA-partisiipista, jonka sisältävä lauseke on kuitenkin huomattavasti yksinkertaisempi kuin lausekkeet B-vaiheen esimerkkivirkkeissä (7) ja (8).

- (9) Ykkösargumentti puuttuu lauseilta 1 ja 4, joissa **näkyvää tekijää** ei ole - joku tuo koiran, joku voi avata ovea, mutta ei se näytä lauseilta. (In2A)

B-vaiheen VA-partisiippiin liittyvät ongelmat johtunevat siis informantin pyrkimyksestä tuottaa pitkiä, runsaasti informaatiota sisältäviä substantiivilausekkeita. Kielenoppimisen vaiheittaisuus ilmenee tämältyypistien havaintojen kautta mielenkiintoisesti: tietynlaiset virheet tietyssä kielenkehityksen vaiheessa kertovat kehityksestä, eivät taantumisesta. Ilmaisun monipuolistuessa predikaatteihin liittyvät ongelmat ovat siis vähentyneet, mutta tekstiin on ilmaantunut inventaarin laajenemisen myötä infinitiivi- ja partisiippirakenteisiin liittyviä ongelmia.

### 3.3.2. Informantin 2 A- ja B-vaihe

Informantin 2 A-vaiheen tenttivastaus sisältää 800 sanetta ja vertailukohteena oleva seitsemän kuukauden kuluttua kirjoitettu tenttivastaus 872 sanetta. Hänen tenttivastauksensa ovat paitsi huomattavasti pidempiä kuin informanttien 1 ja 3, myös huomattavasti pidempiä kuin suomea äidinkielenään puhuvista koostuvan verrokkiryhmän tenttivastaukset (keskimäärin 523 sanetta). Lekseemien suhteellinen määrä informantin 2 teksteissä (n. 120/300 molemmissa vaiheissa) lieneekin tenttivastausten pituuden tuoman toiston vuoksi alhaisempi kuin verrokeilla ja informantilla 1. Erilaisia predikaatteja on kuitenkin informantin 2 tenttivastauksissa huomattavasti enemmän kuin muiden informanttien ja verrokkiryhmän tenteissä, A-vaiheessa 61 ja B-vaiheessa 60. Toisaalta kyseisten tenttien aihepiirit ovat varsin laaja-alaisia.

Informantin 2 A-vaiheen tenttivastauksessa *olla*-verbin osuus predikaateista on vain 23 % (29 kpl). B-vaiheen tentissä *olla*-verbin osuus on kuitenkin noussut 33 %:iin (48 kpl). Odotuksenvastaisesti infinitiivien ja partisiippien määrä on samalla laskenut. A-vaiheessa niitä on

n. 8 kappaletta 300:aa sanaa kohden (abs. 21 kpl) mutta B-vaiheessa vain 5 / 300 (abs. 15 kpl). Informantti käyttää kuitenkin hyvin monipuolisia ja oikein rakennettuja infinitiivi- ja partisiippirakenteita jo A-vaiheen tentissä (ks. asetelma 3).

**ASETELMA 3.** *Informantin 2 A-vaiheen infinitiivi- ja partisiippirakenteita*

<b>A-infinitiivi</b>	1919–1939-luvulla <b>yritettiin luoda</b> liiviläisten identiteetin ja <b>säilyttää</b> heidän unikkaa kieltä tuleville sukupolville ja maailmalle. Toinen <b>tavoite oli järjestää</b> jumalanpalvelun liiviksi.
<b>MA-infinitiivi</b>	Liiviläiset nuoret miehet opiskelivat teologiaa Helsingissä ja <b>olivat</b> opiskelun jälkeen <b>menossa</b> kotialueelle <b>palvelmaan</b> kirkoissa liiviksi.
<b>VA-partisiippi</b>	Inkerikot eli inkeroiset ovat itämerensuomalaiseen <b>pohjoiskieliryhmään kuuluva kansa</b> , joka puhuu inkeroista ja asuu nykyään Leningradin alueen idässä, Viron rajan lähempänä.
<b>TU-partisiippi</b>	Viron kieli oli <b>laajasti käytetty</b> kouluissa ja vuonna 1919 Tarton yliopisto rupesi toimimaan viroksi.

**ASETELMA 4.** *Informantin 2 B-vaiheen infinitiivi- ja partisiippirakenteita*

<b>A-infinitiivi</b>	Myös puhe on eri lauseiden jäsenien <b>mahdollisuuksista osallistua</b> vaihtoon.
<b>MA-infinitiivi</b>	Unkarin australialaiset <b>auttavat ymmärtämään</b> ovatko analysoidut prosessit universaaleja, koska unkari on suomen sukukieli, siis typologisesti läheinen.
<b>VA-partisiippi</b>	Kun <b>dominoivassa kielessä</b> on kehittynyt tarpeeksi upotekielen periaatteita ja puhuja noutaa sen fonologisia, morfologisia, leksikaalisia ja syntaktisia sääntöjä, tapahtuu kääntäminen – upotekielestä tulee <b>dominoiva</b> , siis matriisikieli.
<b>TU-partisiippi</b>	Unkarin australialaiset auttavat ymmärtämään ovatko <b>analysoidut prosessit</b> universaaleja, koska unkari on suomen sukukieli, siis typologisesti läheinen.

Vaikka infinitiivi- ja partisiippirakenteiden suhteellinen määrä on B-vaiheen tentissä pienempi, valikoima on sama kuin vaiheessa A (ks. asetelma 4). Toisin sanoen siitä huolimatta, että informantti 2 osaa käyttää erilaisia infinitiivi- ja partisiippirakenteita oikeaoppisesti, niitä ei esiinny hänen tekstissään kovin taajaan.

Muoto-opillisten virheiden suhteellinen määrä on pysynyt suunnilleen samana; se on laskenut yhdeksästä esiintymästä kahdeksaan 300:aa sanaa kohti (molemmissa 24 kpl). Suurin osa informantin 2 tekemistä virheistä liittyy oikeinkirjoitukseen (esim. äännepituus ja vokaalisointu):

- (10) paikallinen, pakkollinen, oppetti, ääntelliselle, varallista, unikkia, kieltö, sanomalehdessä, vaihoi, jotka, lainettu (In2A)
- (11) kielioppillinen, menetyksesstä, literoinut, haastatelun, tilanetta, perheessaan, sääntöjä, kaksikielisiä, kielenvaihtoon, mitää, ekstalinguistisia (In2B)

Predikaatteihin liittyvät ongelmat ovat kuitenkin vähentyneet. Informantin 2 A-vaiheen tenttivastauksessa predikaateista n. 14 % (18 kpl) on jollain tavalla virheellisiä, B-vaiheen tentissä virheellisten predikaattien osuus on enää n. 5 % (7 kpl). Molemmissa tenttivastauksissa on passiivinmuodostukseen liittyviä ongelmia (*olivat järjestetty, ovat joskus nimitetty, ovat integroitu*). Virheet johtunevat pyrkimyksestä noudattaa kongruenssia ja muistuttavat suomenkielisten virheitä.

- (12) Monesti siellä on säilytetty samat **normit** jotka **olivat** perustettu jo 1800-luvulla. (In2B)

A-vaiheessakin esiintyy myös suomenkielisille kirjoittajille tyypillinen kongruenssivirhe:

- (13) Venäjällä inkeroisia on jäljellä aika vähän, noin tuhat ihmistä ja **ei kaikki** niistä **puhu** enää omaa kieltä. (In2A)

Vaikka informantilla 2 *olla*-verbin suhteellinen osuus on kasvanut ja infinitiivi- ja partisiippirakenteet vähentyneet ja vaikka hän tekee enemmän muoto-opillisia virheitä, hänen tekstinsä vaikuttavat ainakin

käsinkirjoitettuina selvästi parhaimmilta. Mahdollisesti tenttien runsas ja tarkka asiasisältö sekä tekstin erittäin siisti ulkoasu vaikuttavat myös kokonaiskäsitukseen. (Ks. edellä luku 3.1.)

### 3.3.3. Informantin 3 A-, B- ja C-vaihe

Informantilta 3 on edellisistä poiketen tutkimuksessa mukana kolmen eri vaiheen tenttivastauksia. Toisen vaiheen tenttivastaus on kirjoitettu 7 kuukauden kuluttua ja kolmannen vaiheen tentti 14 kuukauden kuluttua tutkimuksen alkamisesta. Informantin 3 ensimmäisen vaiheen (A-vaihe) tenttivastaukset koostuvat kolmesta eri tentistä ja sisältävät yhteensä 1 115 sanetta. Toisen vaiheen (B-vaihe) tentissä saneita on 574 ja viimeisessä (C-vaihe) 792. Sen lisäksi, että tenttivastaukset ovat pidentyneet, myös sanasto on viimeisessä tentissä jo selvästi monipuolisempaa kuin aikaisemmissa: A-vaiheen tenteissä eri lekseemejä on 300:aa sanaa kohden 99, B-vaiheessa 97 ja C-vaiheessa peräti 130.

Myös verbien käyttö on monipuolistunut hieman. A-vaiheen tenteissä on yhteensä 42 eri predikaattiverbiä, kun taas C-vaiheen tentti sisältää yksinään 48 erilaista predikaattiverbiä. A-vaiheen tenteissä *olla*-verbin osuus predikaateista on 36,5 % (49 kpl), vaiheen C tentissä enää 26 % (29 kpl). *Olla*-verbin osuus on siis 14 kuukauden kuluttua 10,5 prosenttiyksikköä vähäisempi. Huomionarvoista on, että B-vaiheen tentissä erilaisten predikaattien määrä on notkahtanut, kun taas *olla*-verbin prosentuaalinen osuus on noussut selvästi lähtötasoon nähden. Erilaisia predikaatteja B-vaiheen tentissä on ainoastaan 15. Vähäisen määrän selittää osittain se, että hieman yli puolet (51 %, 44 kpl) predikaateista on *olla*-verbin eri muotoja.

Ongelmallisten predikaattien osuus väheni 7,5 prosenttiyksikköä B-vaiheeseen mennessä, mutta C-vaiheen tentissä ongelmia on lähes yhtä paljon kuin tutkimuksen alussa (A-vaiheessa predikaattien käyttöön liittyviä ongelmia 16,5 %:ssa (22 kpl), B-vaiheessa 9 %:ssa (8 kpl) ja C-vaiheessa 14,4 %:ssa (16 kpl) predikaateista). Ongelmien laatu paljastaa kuitenkin verbien käytön tarkentuneen tietyiltä osin: kongruens-

siin, subjektin valintaan ja modaalisuuteen liittyviä virheitä ei esiinny C-vaiheen tenttivastauksessa lainkaan ja rektiovirheitäkin on vain yksi. Sekä A- että C-vaiheessa suurin osa ongelmista liittyy verbin valintaan (ks. esimerkit 14 ja 15). (Ks. myös Denisjuk 2010: 68–83.)

- (14) *tapautua* pro *muuttua*:  
 astevaihtelu **tapautuu** heikoksi.  
 pitää: ‘esiintyä, voida esiintyä, pystyä esiintymään tms.’  
 Näistä vain. a, ä ja a **pitivät** toisen tavun vokaaleina, koska kantakie-  
 len tyypillinen sanarakenne oli (c)V(c)Ca/ä/e, Kantakielen sananvar-  
 talo oli siis kaksitavuinen ja vokaaliloppuinen. (In3A)
- (15) *sujua* pro *onnistua*, *pärjätä*  
 Lukijat, jotka ovat parhaimmin testituloksen saajia, eivät **sujunneet**.  
*tuoda* pro *johtaa*  
 Tämä muistuttaa denominatiivista sananmuodostusta, sillä sanaa  
**tuodaan** verbistä välinettä ilmaisevan suffiksi -in kanssa. (In3C)

#### ASETELMA 5. Informantin 3 A-vaiheen infinitiivi- ja partisiippirakenteita

<b>A-infinitiivi</b>	Rekonstruoitu muoto <b>saattaa olla</b> näin: <i>tie-tä-ä : tie-ti</i> < <i>tie-si</i>
<b>E-infinitiivi</b>	Pyöreävokaleissa <i>o</i> esiintyy joko taivutuksen prosessina toisen tavun jäljessä <b>a:n muuttuessa</b> , tai sanojen suffikseissa kuten <i>onneton</i> -sanan jälkitavuisessa <i>-o-</i> .
<b>MA-infinitiivi</b>	Hän <b>pyrki puolustamaan</b> lansimurteet.
<b>VA-partisiippi</b>	<b>On huomioitava</b> kantakielessä ja nykysuomessa yhteisestä piirteestä, että kieltoverbi taipuvat.
<b>NUT-partisiippi</b>	(---) mutta <b>liudentunut</b> [symboli] ei ole säilynyt enää missään kielissä.
<b>TU-partisiippi</b>	<b>Alun perin käytetyt spirantit astevaihtelun heikko- pareina</b> ovat muuttuneet muiksi konsonanteiksi.



Kuten edellisistä esimerkeistä käy ilmi, verbien semanttiset erot ovat usein hyvin hienovaraisia. Koska verbin valitsemista eivät ohjaa samankaltaiset, opeteltavissa olevat säännöt kuin esimerkiksi kongruenssia, voitaneen todeta informantin 3 verbien käytön sekä monipuolistuneen että tarkentuneen 14 kuukauden aikana. Ilmaisun monipuolistumisesta kertonee sekin, että informantti käyttää esimerkkien kaltaisissa tilanteissa merkitykseltään samantyyppistä verbiä sen sijaan, että selittäisi saman asian kiertoilmauksien avulla. Toisin sanoen informantti on tavoitellut synonyymia, mikä myös kertoo kielitaidon monipuolistumisesta.

Informantin 3 A-vaiheen tenttivastauksissa infinitiivi- ja partisiippirakenteita on noin 6 / 300 sanaa (21 kpl) (ks. asetelma 5), siis huomattavasti oletettua (1 / 300) enemmän. B-vaiheessa infinitiivi- ja partisiippirakenteiden suhteellinen määrä on kuitenkin vain 2 / 300 sanaa (4 kpl). C-vaiheen tentissä määrä on kuitenkin noussut 8 esiintymään 300:aa sanaa kohti (20 kpl). Huomionarvoista on myös se, että informantti 3 käyttää infinitiivejä ja partisiippeja A- ja C-vaiheessa hyvin monipuolisesti, vaikkakin B-vaiheen tentti sisältää ainoastaan kolme partisiippia ja yhden, epäidiomaattisesti käytetyn infinitiivin.

Näyttääkin siltä, että B-vaiheen tenttivastaus on verbien osalta yksinkertaisempi erityyppisen sisältönsä vuoksi. Infinitiivisiä rakenteita on vain vähän, ja ne ovat epäidiomaattisia (ks. asetelma 6).

B-vaiheen tentti käsittelee nimittäin saamen sijasysteemin kehitystä, minkä vuoksi tenttivastauksessa on runsaasti verbien *olla* ja *kehittyä* muotoja (esim. *Monikon tunnus on -id, jonka alkuperä oli monikon tunnus -i- + \*-tA, Nämä tunnukset ovat kehittyneet kantakielen \*-mA:n lisäksi elosijasta \*-stA.*).

Informantin 3 tenttivastauksissa on suhteellisen runsas määrä erilaisia muotovirheitä, esim. *huomioittavana, kahten, pluskuvanperfektin, soinnillistä, suljektin* 'subjektin', *tälloin, varhaassa* 'varhaisessa'. Tavurakenteeseen liittyvät muotovirheet (*pluskuvanperfektin*) ja äänneiden /r/ ja /l/ sekoittuminen (*modaariverbien, elosijasta*) ovat tyyppillisiä japanin kielen vaikutukseen liittyviä ilmiöitä. Muotovirheet ovat oletuksenvastaisesti lisääntyneet: A-vaiheen tenteissä virheitä oli hypo-

teesin mukaisesti 5 / 300 sanaa (19 kpl), mutta B-vaiheessa virheitä oli jo 12 / 300 (23 kpl) ja C-vaiheen tentissä peräti 16 / 300 sanaa (41 kpl). Kielitaidon kehitys ja sanavaraston kasvu ovat saattaneet lisätä informantin itseluottamusta kirjoittajana ja sitä kautta mahdollisesti myös kirjoitusnopeutta. Kasvanut kirjoitusnopeus on puolestaan saattanut lisätä huolimattomuusvirheiden määrää. Toisaalta on otettava huomioon myös informanttien yksilölliset tavoitteet: muotovirheiden karsiminen tekstistä ei välttämättä kuulu niihin kielen osa-alueisiin, joita informantti 3 pyrkii kehittämään. Informantti 3 vaikuttaakin kvantitatiivisen analyysin perusteella olevan rohkea kielenkäyttäjä, joka pyrkii ensisijaisesti välittämään asiasisällön.

Informantin 3 B-vaiheen tentissä on useita muista vaiheista poikkeavia lukemia. Kun vertaillaan A- ja C-vaihetta, voidaan havaita selvää kehitystä; vain muotovirheet lisääntyvät täysin odotuksenvastaisesti. (Ks. taulukko 1.)

**TAULUKKO 1.** *Tilastotietoja informantin 3 A-, B- ja C-vaiheen tenteistä. (Tässä informantin 3 kaikki kolme vaihetta, ks. taulukot 2 ja 3, joissa informanttien 1, 2 ja 3 A- ja B-vaiheet esitetään rinnakkain.)*

	A	B	C
lekseemiä / 300 sanaa	99	97	130
virkkeiden keskipituus (sanaa)	13	10	13
predikaatteja	134	86	111
erilaisia finiittiverbejä	42	15	48
<i>olla</i> -verbin %-osuus predikaateista	36,5	51	26
ongelmallisia predikaatteja (%)	16,5	9	14,4
infinitiivejä ja partisiippeja kpl / 300 sanaa	6	2	8
muotovirheet kpl / 300 sanaa	5	12	16

**ASETELMA 6.** *Informantin 3 B-vaiheen infinitiivi- ja partisiippirakenteita*

<b>MA-infinitiivi</b>	Itämerensuomalaisiin kieliin (=ims) verrattuna -ie- on kehittynyt myös pelkkä -i-:stä, jonka jälkeen vokaalit <b>ovat riippumatta</b> .
<b>VA-partisiippi</b>	Aktion esiiviä käytetään yleensä modaariverbien ja <b>aistimista tarkoittavan verbien</b> kanssa, “olla tekemässä” ilmaisussa ja sivulauseen verbinä.
<b>TU-partisiippi</b>	Samoin action lokatiivilläkin on samanaikaista ilmaisu, mutta se toimii mm. <b>estettä ja lakkaamista tarkoitetun verbien</b> kanssa sekä “olla tekeminen” kielteisessä.

**ASETELMA 7.** *Informantin 3 C-vaiheen infinitiivi- ja partisiippirakenteita*

<b>A-infinitiivi</b>	Kaksikielinen 5-vuotias lapsi osaa jo <b>olettaakseni</b> molempia kieliä, eli äidinkieltä ja ns. perhekieltä, jos sanotaan Suomessa asuvasta ja syntymästä 5-vuotiaasta lapsesta.
<b>MA-infinitiivi</b>	Arvaan, että aikuinen <b>on selittänyt</b> aikaisemmin junasta ‘peräkkäin’-sanan <b>käyttämällä</b> , ja lapsi muistaa, että ‘peräkkäin’ ja ‘juna’ liittyvät jotenkin.
<b>VA-partisiippi</b>	Nämä seikat muistuttavat mielestäni <b>vieraita kieliä</b> <b>opiskelevien ulkomaalaisten</b> opetusta.
<b>TU-partisiippi</b>	CDS on hoivakielen termin lyhennys, jolla tarkoitetaan suomeksi <b>lapselle suunnattua kieltä</b> .
<b>nippu (t)Tavissa</b>	Sen mukaan vertaan yksikieliseen lapseen (tässä tapauksessa suomalaisen lapseen), olisiko heillä suuriero suomen kielen omaksumisessa, ja jos olisi niin miten se <b>on havaittavissa</b> .
<b>agenttiparti- siippi</b>	<b>Mäkisen luomat tehtävätyypit</b> ovat yhteensä 12: riimitehtävä, sanojen tavun poistuminen ja lisääminen, sanojen pituudesta, sanojen tavuttaminen, sanojen alkutavun ja lopputavun nimeäminen, sanojen alkutavun ja lopputavun ääntäminen. jne.

### 3.4. Verrokkit

Verrokkeina on kolme suomenkielistä suomen kielen opiskelijaa. Heiltä kullakin on yksi tenttivastaus tutkittavana.

Infinitiivi- ja partisiippiarakenteita verrokkiryhmällä on huomattavasti enemmän (23 kpl / 300 sanaa) kuin informanteilla. Kaikilla verrokeilla on erittäin monipuolisesti erilaisia infiniittisiä rakenteita (esimerkki 16). Mukana on muun muassa pitkiä määräiteketjuja, jotka eivät informanteilta vielä onnistu; muutamissa taas infiniitivirakenne määrittää toista infiniitiiviä.

- (16) Siksi onkin tavallista, että erityisesti **suomenruotsalaisiksi itsensä luokittelevat** kaksikieliset suomalaiset lähettävät lapsensa ruotsinkieliseen kouluun, vaikka perheen oma hallitseva kieli olisikin suomi. (Verrokki 1)

Sisältösanojen ja muotosanojen välisen eron ymmärtämistä **voi** myös **helpottaa jättämällä** lauseesta pois joko sisältösanan tai muotosanan. (Verrokki 2)

Lapsella on tällöin jo tietyt kielelliset ja mentaaliset mallit, joita hän **voi hyödyntää oppiessaan** toista kieltään. (Verrokki 3)

Toisaalta verrokkiryhmällä myös *olla*-verbin osuus on suurempi kuin informanteilla. Verbien käyttöön liittyviä ongelmia verrokkiryhmäläisillä ei juuri esiinny (yhdellä verrokilla 2 kpl (3,7 %), kahdella muulla predikaatteihin liittyviä ongelmia ei esiinny lainkaan), kun taas informanteilla virheellisten predikaattien osuus on keskimäärin 12,4 %. Koska kaikilla informanteilla ongelmallisten predikaattien osuus myös laski tutkimuksen edessä, näyttää siltä, että verbien käyttöä ohjaavat opeteltavissa olevat säännöt, jolloin virheet vähenevät kielitaidon kohentuessa.

Muoto-opillisia virheitä esiintyy kuitenkin jonkin verran (3 / 300) myös verrokkiryhmäläisillä. Verrokilla 1 muotovirheitä on n. 6 / 300 sanaa (10 kpl), mikä on suunnilleen saman verran kuin informantin 3 A-vaiheen tenttivastauksissa (5/300, 19 kpl) ja informantin 1 molempien vaiheiden tenttivastauksissa (7/300) (A 9 kpl ja B 17 kpl). Muotovirheiden syntyyn vaikuttavat seikat, kuten huolellisuus, stressi tai kiire,

näyttävätkin olevan ainakin jossain määrin kielitaidon ulkoisia tekijöitä. Verrokkien muotovirheitä tarkastelemalla virheiden tilapäisyys tulee selvemmin ilmi kuin informanttien tekemistä muotovirheistä, koska virheitä ei voi selittää puutteellisella tiedolla (ks. esimerkki 17).

- (17) omankielistäopetusta; ruotsinkieli; sosiaalisen verkostoitumisen painostus; Tämä vaihe on luovaa ja sääntöjä yliyleistävä tuottoisaa (Verrokki 1)

sillä tarkoitetaan toisensa poissulkevaa vastakohtaisuutta merkitysmuutoksesta (Verrokki 2)

### 3.5. Kvantitatiiviset tutkimustulokset vertailtavina

Taulukoissa 2 ja 3 on esitetty luvuissa 3.3 ja 3.4 käsitellyt ilmiöt kvantitatiivisesti. Verrokkiryhmän lekseemimäärä ei eroa paljoakaan informantin 1 A-vaiheen tenttien lekseemimäärästä, kun taas informantilla 3 lekseemimäärät ovat sekä A- että B-vaiheen tenteissä melko alhaiset.

Taulukon 3 yksityiskohtaiset lukemat osoittavat informanttien keskinäiset suhteet ja suhteet verrokkeihin. Huomionarvoista on, kuinka vähän eri predikaattiverbejä verrokeilla on. Tätä lukemaa selittää se, että verrokeilla on runsaasti erilaisia infinitiivejä ja partisippeja, minkä vuoksi huomattava osa predikaattiverbeistä on modaaliverbejä ja *olla*-verbin eri muotoja. Informantilla 2 taas on erityisen vähän *olla*-verbejä A-vaiheen tentissä. Niiden lisääntyminen ei siis hänen tapauksessaan suinkaan ole merkki kielen yksipuolisuudesta, vaan määrän normalisoinnista lähelle äidinkielisten tasoa.

### 3.6. Kvalitatiivinen katsaus informanttien tuotoksiin

#### 3.6.1. Tenttivastausten verbit ja niiden erityisluonne

Kaikkien informanttien tenttivastauksissa esiintyy odotuksenmukaisesti runsaasti ilmiöiden kuvaamiseen ja selittämiseen liittyviä verbejä esim. *aiheutua, esiintyä, hävitä, ilmestyä, johtua, kehittyä, kuulua, liittyä,*

**TAULUKKO 2.** *Tenttivastausten sane- ja lekseemimäärät sekä virkkeiden keskipituus informanteilla 1–3 ja verrokkiryhmällä*

	Informantti 1		Informantti 2		Informantti 3		Verrokki-ryhmä
	A	B	A	B	A	B	
tenttejä kpl	2	1	1	1	3	1	1
sanoja	413	694	800	872	1115	570	523
lekseemejä	206	326	327	353	367	185	261
lekseemiä / 300 sanaa	149	141	123	121	99	97	153
virkkeiden keskipituus (sanaa)	13	16	11	14	13	10	12

**TAULUKKO 3.** *Tenttivastausten predikaatit, nominaalimuodot ja muoto- sekä kirjoitusvirheet informanteilla 1–3 ja verrokkiryhmällä*

	Informantti 1		Informantti 2		Informantti 3		Verrokki-ryhmä
	A	B	A	B	A	B	
predikaatteja	59	87	125	146	134	86	63
erilaisia finiittiverbejä	20	31	61	60	42	15	25
<i>olla</i> -verbin %-osuus predikaateista	52,5	34,5	23,2	32,9	36,5	51	37,1
ongelmallisia predikaatteja (%)	15,3	8	14,4	4,8	16,5	9	1,2
infinitiivejä ja partisiippeja kpl / 300 sanaa	8	12	8	5	6	2	23
muotovirheet kpl / 300 sanaa	7	7	9	8	5	10	3

muodostaa, muuttua, näyttää, puuttua, riippua, säilyä, tapahtua, tarkoittaa, vaikuttaa. Samoin tutkimukseen ja sen esittelyyn liittyviä verbejä on kaikkien informanttien teksteissä runsaasti, esim. *analysoida, esitellä, esittää, kuvata, käsitellä, käyttää, luokitella, nimittää, raportoida, ryhmitellä, säilyttää, tarkastella, verrata, viitata, yhdistää*. Myös modaaliverbejä (*haluta, osata, pitää, pystyä, saada, saattaa, tahtoa, voida*) esiintyy kaikkien informanttien teksteissä (vrt. Niiranen 2010: 17–21). Informanttien tenttivastausten viisi yleisintä verbiä ovat *voida, tapahtua, puhua, kehittyä* ja *esiintyä*. (Vrt. verbien semanttinen luokittelu Pajunen 2001.)

Tenteissä esiintyy myös jonkin verran esittelyyn ja puhumiseen yleisesti liittyviä verbejä. Esimerkiksi informantin 1 toisen vaiheen tentissä tällaisia verbejä on neljä erilaista (*kertoa, mainita, puhua, todeta*). Lisäksi tenttivastauksissa on lingvistiikkaan liittyviä verbejä (*kongruoida, litteroida, rekonstruoida, sulautua, vaihdella, ääntää*) sekä muutamia ihmisen mentaaliin toimintoihin liittyviä verbejä (*ajatella, muistaa, tietää, tuntea, ymmärtää*). Jälkimmäisistä muutama esiintyy myös yksikön ensimmäisessä persoonassa informantin 1 tenttivastauksissa (18–20).

- (18) ‘Nimi’ on hankala - minun mielestäni se noudattaa kaikki sääntöjä, mutta **tiedän** kuitenkin, että on lainaus. (In1A)
- (19) En **tunne** sanaa “kalvo”. (In1A)
- (20) Pahoittelen, mutta **en tiennyt** tästä artikkelista - minulla oli pelkääntään lehtileikkeet. (In1B)

Arkisia, merkitykseltään hyvin konkreettisia verbejä (*laulaa, nukkua, opettaa, vieraila*) on lähinnä informantin 2 A-vaiheen tentissä, jossa informantti kuvaa pienten kieliyhteisöjen asemaa ja toimintaa sekä erilaisia hankkeita, joita on kielten säilyttämiseksi toteutettu. Tuttuudestaan huolimatta konkreettiset verbit ovat tenttivastauksessa marginaalisia, ja niiden esiintyminen näyttäisikin riippuvan eniten tenttikäsymysten luonteesta ja tenttien aihepiireistä.

### 3.6.2. Informanttien virkkeiden esittely

Vaikka kvantitatiiviset tulokset eivät kerro yksiselitteisesti edistyksestä, näyttää siltä, että holistisen arvion (ks. mm. Tarnanen 2002: 88, 265) pohjalta katsottuna kahden tutkitun kauden eri tekstit ovat hyvin erilaisia: B-vaiheen tekstit ovat ehdottomasti ymmärrettävämpiä kuin A-vaiheen. Käsinkirjoitetusta tekstistä saa helposti vielä positiivisemmän mielikuvan kuin tässä esillä olevista digitalisoiduista esimerkeistä.

Seuraavassa on kultakin informantilta yksi virke tai lyhyt kappale sekä A- että B-vaiheen tentistä. Näiden avulla haluamme osoittaa, että siitä huolimatta, että monet yksittäiset laskettavat ilmiöt eivät 7 kk:n opiskelun aikana kovastikaan muuttuneet, teksteistä alkaa B-vaiheessa löytyä hyviä kokonaisuuksia. Toki A-vaiheessakin niitä esiintyy, mutta ei niin monipuolisesti.

- (21) Sajnovics halusi osoittaa unkarin ja saamen kielten välisten suhteet vääräksi, sen takia että hän piti unkarin kieli saamen niin hyvin korkeampana. (In1A)

Informantin 1 esimerkissä (21) lausekkeet eivät toimi, koska määritteiden sijamuodot määräytyvät vasemmalta eivätkä oikealta. Myös komparatiivin vahvikesana on virheellinen eikä objekti ole oikeassa sijassa.

- (22) Yksi suomen monista sijamuodoista on partitiivi, jolla voi usein erottaa määräisen tilanteen epämääräisestä - esim. kun monikon nominatiivi toimii määräisenä ja monikon partitiivi epämääräisenä. Partitiivi löytyykin suomen sukukielistä, esim. virosta. Toiseksi, suomesta puuttuu kieliopillinen suku. (In1B)

Informantin 1 B-vaiheen tekstikokonaisuus on onnistunut ja sen rakenne on korrekti. Ainoa ongelma sujuvassa katkelmassa on, että liitepartikkelin on ymmärretty myös verbiin liittyvänä ilmaisevan samaa kuin sanan *myös*. Myös natiivien teksteissä *-kin*-liitepartikkelit saattavat olla lauseessa epäonnistuneessa kohdassa.

- (23) Syinä oli pakkollinen venäläistyminen ja uskonnon kieltö. (In2A)



A-vaiheessa informantilla 2 on kvantiteetti- ja vokaaliharmoniaongelmia, joskin *kieltö*-tapauksessa varmasti opetukseen ja loogiseen päätteelyyn perustuva ongelma, vrt. etuvokaalinen *kieltäminen*. Opetuksessa ei todennäköisesti ole selitetty tiettyjen johdinten erityisasemaa vokaaliharmoniassa. Lauseen alkuosa on tulkittavissa kahdella tavalla: a) ‘venäläistyminen pakon edessä’ tai b) ‘pakkovenäläistäminen’. a-tulkinnassa virhe olisi vähäisempi, vain määräite on hieman epäonnistunut; vakavammaksi ongelmaksi voitaisiin laskea b-tulkinnan *venäläistyminen*-substantiivin käyttö substantiivin *venäläistäminen* sijaan. Tällöin olisivat agenttisuhteet menneet sekaisin.

- (24) Kovácsin väitöskirja käsittelee koodinvaihtoa ja kielenvaihtoa. Tutkimuksen informantit ovat kaksikielisiä, kaikki puhuvat englantia ja asuvat Australiassa, jossa englanti on enemmistökieli. Toisaalta informanteilla on joko suomalainen tai unkarilainen tausta. (In2B)

Informantti 2 pystyy kirjoittamaan B-vaiheessa idiomaattista ja erinomaista suomenkielistä asiaproosaa. Tenttivastausten käyttöä tutkimusaineistona ei aina ole pidetty hyvänä, koska on esitetty myös näkemyksiä, että opiskelijat lukisivat ulkoa katkelmia tenttikirjoista ja tällöin kieli ei olisi heidän omaansa. Tässä esimerkissä tenttikirja on kuitenkin kirjoitettu englanniksi, joten ulkoaoppimisesta ei voi olla kysymys.

- (25) Samoin koskee myös pronominin että kieltoverbin vartalo on olemassa etymologiseltaan hyvin vanhasta ajasta. (In3A)

Informantin 3 virke (25) ei ole jäsennettävissä yksiselitteisesti. Siinä on kyllä yksittäisiä korrekkeja lausekkeita, mutta kokonaisuus jää hahmottomaksi.

- (26) Lokatiivin tunnus on yksikössä -s, joka on kehittynyt kantakielen latiivista \*-s ja tulosijasta \*-tA. Attribuutteina toimivat saamessa pronominit, numeraalit, adjektiivit ja infiniittiset verbit. Näistä numeraalit eivät kongruoi pääsanana kanssa vaan ne ovat genetiivi-akkusatiivissa. (In3B)

Informantin 3 B-vaiheen tentistä on löydettävissä jo aivan korrektejakin kokonaisuuksia (26).

Kaiken kaikkiaan tämän tutkimuksen tulokset viittaavat siihen, että edistynyt akateeminen oppija ei enää yksityiskohdissa juuri poikkea natiiveista kielenkäytössään ja epätarkkuudetkin ovat samankaltaisia kuin natiivien epätarkkuudet. Tässä artikkelissa pyritäänkin tarkastelemaan, poikkeaako kielenoppijoiden kieli joiltakin tietyiltä frekvensseiltään natiivien kielestä. Lisäksi tosin halutaan myös nähdä, tapahtuuko joissakin korrektiuteen liittyvissä ilmiöissä muutosta 7 kk:n intensiivisen opiskelun jälkeen.

## 4. Pohdintaa

### 4.1. Tenttivastaukset tutkimusmateriaalina

Tenttikysymyksillä on kiistatta vaikutus tenttivastausten sisältöön. Esimerkiksi tenttivastausten pituus riippuu pitkälti tenttikysymysten luonteesta. Jotkin tentit saattavat koostua lyhyistä selitystehtävistä ja valmiista esimerkivirikkeistä ja toiset pitkiä esseevastauksia edellyttävistä kysymyksistä. Toisissa tenttitilaisuuksissa saattaa myös olla huomattavasti enemmän aikaa kuin toisissa kysymymäärään suhteutettuna. Näin ollen tenttivastauksen pituutta ei voida pitää luotettavana kielitaidon tai tenttivastauksen laadun mittarina. Myös verbien tempuksen ja moduksen valinta riippuu pitkälle siitä, onko tekstissä tarkoitus esitellä esimerkiksi jonkin ilmiön toimintaa, syntyä vai sen meneillään olevaa muutosprosessia. Esimerkiksi informantin 3 A-vaiheen tenttivastauksissa 63 % predikaateista on preesensissä ja 16 % perfektissä ja C-vaiheen tentissä preesensissä on peräti 83 % ja perfektissä ainoastaan 4 % predikaateista. Tempusten ja modusten jakauma tenttivastauksissa ei siis kerro oppilaan kielenkehityksestä. Infinitiivi- ja partisiippirakenteiden suhteellinen määrä ei kuitenkaan riipu samalla tavoin tenttikysymysten laadusta. Informanttien A-vaiheen tenteissä infinitiivi- ja partisiippirakenteita on keskimäärin 8 kappaletta ja B-vaiheen tenteissä runsaat 6 kappaletta 300:aa sanaa kohden. Edistyneet oppilaat käyttävät tosin sanoen jo alku-

vaiheessa oletusta enemmän infinitiivi- ja partisiippirakenteita. Suomea äidinkielenään puhuvien verrokkiryhmällä infinitiivejä ja partisiippeja on kuitenkin vielä huomattavasti enemmän: 23 kappaletta 300:aa sanaa kohden.

Lisäksi on hyvä ottaa huomioon, että informantit ovat paitsi voineet suorittaa tenttejä keskenään eri järjestyksessä myös valita opintokokonaisuuksiinsa keskenään hyvin erilaisia opintojaksoja. Näin ollen samankaan luentokurssin tenttejä ei ole pitkittäistutkimuksen kannalta mielekäästä verrata toisiinsa, jos ne on suoritettu opintojen eri vaiheessa.

## 4.2. Ilmiöiden keskinäiset suhteet

Usein eri ilmiöt limittyvät toisiinsa niin, että yhtä ilmiötä tutkimalla saadun tuloksen perusteella voidaan päätellä jotakin myös jonkin toisen ilmiön esiintymisestä. Näin on esimerkiksi tässä tutkimuksessa tarkastellun *olla*-verbin prosentuaalisen osuuden merkitystä arvioitaessa. Runsaasti *olla*-verbejä sisältävät tekstit eivät välttämättä kerrokaan tekstin olevan vähemmän tyypillistä suomea, vaan *olla*-verbien korkea prosentuaalinen osuus voi johtua esimerkiksi tentin luonteen vaatimasta lauserakenteesta, kuten informantin 3 B-vaiheen tentissä. Niina Salmi onkin pro gradu -tutkielmassaan tutkinut viiden LAS2-informantin kielenkäyttöä, erityisesti heidän lausetyyppejään ja virkerakenteitaan. Salmen laskelmista voi nähdä, että informantin 3 nimenomaan käsillä olevassakin artikkelissa tutkitun B-vaiheen tekstissä on kopulalauseita hyvin runsaasti (Salmi 2010: 79). Kun tarkastellaan tekstin sisältöä, voidaan todeta, että kielihistoriaa ja saamen kielen kehitystä esittelevä teksti sisältää ymmärrettävästi paljon toteavia kopulalauseita.

Suomalaisen verrokkiryhmän *olla*-verbien melko korkea prosentuaalinen osuus ei kuitenkaan kerro kopulalauseiden suuresta määrästä vaan on yhteydessä infinitiivi- ja partisiippi-ilmausten runsauteen: juuri verbi-liittoja ja erilaisia infinitiivi- ja partisiippirakenteita sisältävät tekstit vaativat usein ns. värittömiä verbejä, kuten verbiä *olla*. *Olla*-verbillä on siis yhteys paitsi kopulalauseisiin myös infinitiivi- ja partisiippirakenteisiin.

Salmen (2010) tutkielma vahvistaa myös usein esitettyä hypoteesia, jonka mukaan kompleksisia tekstejä rakentavat päätyvät useampiin virheisiin kuin lyhyempiä virkkeitä rakentavat. Tämä on näennäisesti aivan itsestään selvää. Nyt ei kuitenkaan edistyneiden oppijoiden teksteistä, jotka kaiken kaikkiaan ovat melko hyvää suomen kieltä, mitenkään vain lukemalla huomaa, että korrektius tosiaan saattaa korreloida esimerkiksi sivulauseiden määrän kanssa, koska sivulauseita esiintyy kaikkien informanttien teksteissä suhteellisen paljon. Sekä varman päälle ja varovasti kirjoittavien että rohkeasti monipuolista kieltä käyttävien tekstit vaikuttavat ensi lukemalta vaihtelevan laadultaan satunnaisesti, mutta tarkkojen täsmälaskelmien jälkeen kyse ei näytäkään olevan vapaasta variaatiosta vaan tiettyjen ilmiöiden keskinäisistä suhteista. (Ks. Salmi 2010: 97.)

## 5. Muutos

Tässä tapaustutkimuksessa olemme tarkastelleet edistyneiden oppijoiden kehittymisen mahdollisuutta. Tutkimuksen tarkasteluajankohtien väli on vain n. 7 kuukautta. Pitkälle edistyneen kielenoppijan kielitaito voikin näyttäytyä hyvin vaihtelevana eri tarkasteluajankohtina (vrt. Salmi 2010: 56, 60). Muutoksiin saattavat näin korkealla tasolla vaikuttaa enemmän esimerkiksi tentin luonne kuin se, että on kulunut aikaa A-vaiheesta B-vaiheeseen. Tietty muutoksen suunta on kuitenkin nähtävissä. Predikaatteihin liittyvät ongelmat vähenevät, ja tästä aiheutuu se, että lauserakenteet käyvät selväpiirteisemmiksi ja ymmärrettävämmiksi.

Monissa toisen kielen omaksumisen tutkimuksissa on tultu siihen tulokseen – joskin näkemystä on myös kritisoitu (ks. Ellis 1994: 109–110) –, että kielen omaksumisessa tietyt ilmiöt yleensä omaksutaan tiettyjä ilmiöitä aikaisemmin ja tiettyjä myöhemmin. Oppimisjärjestys olisi siis tämän mukaan eri oppijoilla jossain määrin samanlainen, vaikka kunkin hetkinen keskinäinen kielitaito olisikin hyvin erilainen. (Ks. esim. Goldschneider & DeKeyser 2005: 61–64; Pienemann 2005: 41.) Tätä ilmiötä pyrkii taulukko 4 osaltaan havainnollistamaan.

**TAULUKKO 4.** Tutkittavissa piirteissä tapahtunut muutos 7 kk:n opintojen jälkeen informanteilla 1–3

Muutokset	Informantti 1	Informantti 2	Informantti 3
virkkeiden pituus (sanaa)	+3	+3	-3
<i>olla</i> -verbien määrä (prosenttiyksikköä)	-18	+9,7	+14,5
virheellisiä predikaatteja (prosenttiyksikköä)	-7,3	-9,6	-7,5
muoto- tai kirjoitusvirheitä (kpl / 300 sanaa)	0	-1	+7
infinitiivejä ja partisiippeja (kpl / 300 sanaa)	+4	-3	-4

Ensi silmäykseltä tulokset tosin näyttävät hajoavan aivan eri suuntiin, mutta kun taulukkoa verrataan edellä oleviin taulukoihin 2 ja 3, huomataan arvojen pysyvän tietyllä vaihteluvälillä. Toisaalta kielen käyttöön pohjaavissa malleissa (engl. *Usage-Based Models*) lähestytään kielen oppimista muun muassa suhteessa ympäristön antamaan kielelliseen kokemukseen (ks. esim. Kemmer & Barlow 2000) eikä pohjata tiettyyn ennalta ajateltuun järjestykseen.

Kuten edellä todettiin, vain virheellisten predikaattien määrä muuttui kaikilla selvästi samansuuntaisesti ja odotuksenmukaisesti. Seuraavassa pohditaan luvussa 3.1 esitettyjen hypoteesien toteutumista (ks. myös taulukot 2 ja 3).

- 1) Hypoteesi *olla*-verbien määrästä ei toteutunut: korkein prosenttiosuus tutkimuksessa ylipäänsä on vain n. 52 %. (Ei siis 80 %, kuten oletettiin, mikä olisikin tehnyt tekstit erittäin yksitoikkoisiksi.) Arviona oli, että lopussa prosenttiosuus olisi n. 30 %-yksikköä pienempi; vaihteluväli onkin kaiken kaikkiaan n. 30 %-yksikköä – muutos vain ei ole selväpiirteisesti yksisuuntainen. Verbivalikoima on informantilla 2 monipuolinen alusta alkaen, informantilla 3 vaihtelevin ja informantilla 1 suunnilleen sama kuin verrokki-

ryhmillä. Verbivalikoima on vahvasti sidoksissa tentin aihepiiriin (ks. edellä luku 3.6.1). Teksteistä saattaa puuttua aivan frekventtejäkin verbejä, mutta toisaalta niissä on tiettyjä kielitieteen alaan kuuluvia verbejä, jotka eivät esiinny Suomen kielen taajuussanas-  
tossa (TS). Teksteistä saattaa myös olla niukasti joitakin muoto-  
ryhmiä verbivalikoiman myötä: jos tekstissä ei ole modaaliverbejä,  
ei tiettyjä A-infinitiivirakenteitakaan voi tekstissä olla.

- 2) Muotovirheitä on oletetun vähän (alus-  
sa 5/300) vain informantin 3 A-vaiheen tekstissä; kaikissa muissa teksteissä niitä on enem-  
män. Tässä lienee kyse tenttivastausten spontaaniudesta, joka  
vaatii asiatiitoon keskittymistä ja sallinee lipsahdukset ja esimer-  
kiksi vokaaliharmonia- ja kvantiteettivirheet. Huomionarvoista  
on, että lipsahdus- tai muotovirheitä on myös äidinkielisillä ver-  
rokeilla (3 / 300).
- 3) Ongelmia predikaattiverbeissä on sekä A-vaiheessa (14, 4–16,5%)  
että B-vaiheessa (4,8–9 %) oletettua vähemmän.
- 4) Partisiippi- ja infinitiivirakenteita on oletettua enemmän jo heti  
alkuvaiheessa: ensimmäisissä tenteissä niitä on 6–8 / 300, myö-  
hemmin 5–12 / 300. Niitä on kuitenkin huomattavasti vähem-  
män kuin äidinkielisellä verrokkiryhmällä (n = 23).

Olemme yhdistäneet kvantitatiiviset keinot taustatietojen antamaan  
informaatioon sekä tekstin lähilukuun ja näin vertailemme muutosta  
informanttien kesken. Luvussa 2 esitimme 4 tutkimuskysymystä. Seuraa-  
vassa kokoamme tutkimuksemme perusteella vastauksia kysymyksiin.

Ensimmäiseksi kysyimme, onko havaittava muutos mahdollinen  
myös edistyneiden oppijoiden tasolla ja kuinka rajallista kielitaidon  
kehittyminen tässä vaiheessa on? Edistyneiden oppijoiden kieli kehit-  
tyy erityisesti, jos he koko ajan intensiivisesti opiskelevat kieltä. Tässä  
tutkimuksessa informantit 1 ja 2 olivat opiskelleet suomea ennen tut-  
kimuksen alkua Suomen ulkopuolella, mutta juuri tarkastelun alkaessa  
he aloittivat opintonsa Suomessa ja ovat omaksuneet koko tutkimuksen  
ajan uudenlaista diskurssia. Tämä näyttää kiihdyttäneen oppimista.

Toiseksi kysymme, ovatko muutostendenssit nähtävissä havainnollisesti ja verrattavissa keskenään, jos ja kun lähtötilanne vaihtelee eri informanttien välillä. Muutostendenssejä voidaan havaita ja verrata, mutta tällöin myös absoluuttiset arvot on otettava huomioon, jotta nähdään edistyneen oppijan tiettyjen mitattujen arvojen todennäköinen vaihteluväli ja myös sen suhde äidinkielisten puhujien keskiarvoarvoihin. Vaihteluväli on tärkeä ottaa huomioon erityisesti, koska informanttien lähtötilanteet ovat erilaiset, mutta myös siksi, että tehtäviin liittyvät taustamuuttujat vaikuttavat suuresti tuloksiin. Lähtötilanteesta on tietyt minimivaatimukset: mukana on vain maisteriohjelmaan valittuja henkilöitä, joilla on tietty kielitaidon taso, mutta vain alin taitotaso on määriteltävissä, ylärajaa ei ole. Tällöin lähtötilanne on kaikilla hyvin erilainen, mutta jokaisella on jo opinnot aloittaessaan laaja sanavarasto ja monipuolinen suomen kielen rakenteen tuntemus.

Kolmanneksi esitimme kysymyksen siitä, voidaanko muutosaikaa käyttää yhdistävänä tekijänä lähtötilanteesta riippumatta. Jos valitut indikaattorit ovat kielitaitoa onnistuneesti mittaavia, muutos aika näyttää olevan mahdollinen yhdistävä tekijä. Tässä tutkimuksessa erityisesti predikaattiin liittyvät ongelmat ja erilaiset kirjoitusvirheet näyttivät olevan indikaattoreita, jotka eivät niinkään liity yksittäisiin tehtäviin vaan nimenomaan kirjoittajan taitojen muutoksiin.

Lopuksi esitimme vielä laajempia korpuksia koskevan kysymyksen siitä, miten eri informanttien aineistoja voisi sitoa toisiinsa, jotta voitaisiin saada aikaan määrällisesti merkittäviä otantoja. Kun tehtäisiin useampia tutkimuksia, jotka olisivat tässä tehdyn pilotti- ja tapaustutkimuksen kaltaisia, saataisiin todennäköisesti esille luotettavia indikaattoreita. Näitä eri indikaattoreita voitaisiin tutkia yhtäaikaaisesti useilta eri informanteilta, joiden aineistot olisi kerätty tasavälein. Näin saataisiin riittävän suuri otanta, joka kertoisi koko informanttiryhmän kielitaidon kehityksestä. (Vrt. tässä käytetyn LAS2-aineiston korpuksesta tehty pilotti, jossa tutkitaan eksistentiaali- ja intransitiivilauseiden esiintymien frekvenssiä, Ivaska & Siitonen (tulossa).) Virheellisten predikaattien tapauksessa sitominen onnistuu jo kolmenkin informantin ryhmällä.

Virheellisten predikaattien määrä on vähentynyt yhteensä kaikilla informanteilla tilastollisesti merkitsevästi (p-arvo 0.002174); yksittäisen informantin muutoksesta ei sen sijaan voi osoittaa tilastollista merkitsevyyttä tapausten vähäisyyden vuoksi.

## 6. Lopuksi

Olemme esittäneet tiettyjä ilmiöitä kolmen informanttimme kielen muutoksista 7 kk:n aikana. Aikaväli on lyhyt, mutta toisaalta kielenoppijat ovat aivan uudessa vaiheessa opiskelussaan, erityisesti informantit 1 ja 2. Tärkeä kysymys kuuluu, ovatko valitsemamme seikat ollenkaan sellaisia ilmiöitä tai piirteitä, jotka voidaan laskea todellisiksi kielen kehitystä ilmaiseviksi indikaattoreiksi.

Edellä mainittu Salmen tutkimus viiden kielenkäyttäjän tiettyjen piirteiden variaatiosta tuo esiin jossain määrin samankaltaisia seikkoja kuin omamme. Hän on valinnut indikaattoreiksi muun muassa sivulauseiden määrän ja eri lausetyyppien esiintymät. Näyttääkin siltä, että virkerakenne voisi olla toimiva kielitaidon indikaattori, kun siitä saadut tulokset suhteutetaan kunkin tarkastelukohdan erityisominaisuuksiin. (Salmi 2010: 100–103.)

Tilastolliset seikat vaikuttavat siihen, miltä kieli kuulostaa ja vaikuttaa. Tilastollisiin seikkoihin on vaikea päästä käsiksi ilman korpusta. Maisa Martin on kuitenkin korostanut korpusten rakentajien tarvitsemaa toisen kielen oppimisen asiantuntemusta ja käsitteellisen määrittelyn tärkeyttä, jotta korpusten käyttäjät tietäisivät, minkä tiedon pohjalta ja millä perustein kerääminen on tehty (Martin 2010). Validilla tavalla tehty tilastollinen analyysi ja siihen liittyvä holistinen arvio tukevat toisiaan. Molemmat menetelmät näyttävät johtavan suunnilleen samaan lopputulokseen, vaikka tutkimusprosessit poikkeavatkin toisistaan huomattavasti.

Jos noin puolen vuoden välein poimittujen otosten tulokset poikkeavat toisistaan merkittävästi, saattaa menossa olla opitun uudelleenjärjestyminen. (Ks. Ellis 1985: 95–96; Housen 2002: 96–97; Ivaska 2009: 111,



118–119, 122.) Eroihin saattavat olla vaikuttamassa kuitenkin tekstin luonteesta tai sisällön leksikaalisista vaatimuksista johtuvat seikat – tällöin kyse onkin samankaltaisista seikoista kuin äidinkielisillä puhujilla eikä suinkaan kielitaidon rajallisuudesta.

Korpustutkimuksessakaan ei ole syytä luottaa sokeasti kvantitatiivisiin tuloksiin, vaan tutkijan pitää palata tekstiin, miettiä sen semanttis-pragmaattisia vaatimuksia ja niiden kytkeytymistä syntaksiin ja morfologiaan ja vasta sen jälkeen tehdä arvioita – myös tekstilajit ja aihepiirit vaikuttavat osaltaan lopputulokseen. Kvantitatiivisen analyysin jälkeen on siis syytä tutustua teksteihin uudelleen, lukea, ymmärtää ja tulkita ne ja tehdä vasta sitten lopullinen analyysi ja vertailu.

Peruskysymykset, joihin tulevaisuudessa haluaisimme saada vastauksen, ovat seuraavat:

- a) Onko vielä pitkälle edistyneiden oppijoiden tasolla tiettyjä alkeellisia ongelmia, jotka saattaisivat johtua seurantaan varhaisemman opetuksen painotuksista?
- b) Mitkä sitten ovat sellaisia viimeisiä ilmiöitä, joiden vuoksi puhutaan miltei natiivin tasolla olevista puhujista eikä natiivin tasolla olevista puhujista?
- c) Miten erilaiset tiettyihin tyyppillisyyksifrekvensseihin liittyvät ilmiöt, esimerkiksi tiettyjen lausetyyppien esiintyminen, passiivin käyttö jne., poikkeavat oppijoilla äidinkielisten ilmiöistä? (Ks. myös Ivaska & Siitonen 2010.)

Nämä kysymykset redusoituvat seuraavaksi pääkysymykseksi: onko edistyneidenkin oppijoiden suomi kokonaisrakenteeltaan jollakin lailla erilaista kuin äidinkielisten suomi? Toivomme, että tämä tutkimus on luonut pohjaa ja antanut ideoita seuraaviin tutkimuksiin, joissa etsitään vastauksia edellä oleviin kysymyksiin.

## **Kiitokset**

Kiitämme Turun Yliopistosäätiötä LAS2-hankkeen saamasta tuesta.

## **Aineslähteet**

LAS2 = Edistyneiden suomenoppijoiden hankkeen korpus, Lauseopin arkisto, suomen kieli, kieli- ja käännöstieteiden laitos, Turun yliopisto.

## **LAS2-tenttivastauskoodit**

In1 = Informantti 1

In1A = Informantin 1 A-vaihe:

las2-17tt01te02

las2-17tt01te04

In1B = Informantin 1 B-vaihe:

las2-17tt01te04

In2 = Informantti 2

In2A = Informantin 2 A-vaihe:

las2-19tt01te06

In2B = Informantin 2 B-vaihe:

las2-19tt01te05

In3 = Informantti 3

In3A = Informantin 3 A-vaihe:

las2-12tt01te01lo21

las2tt01te02lo21

las2-12tt01te03lo22

In3B = Informantin 3 B-vaihe:

las2-12tt01te06lo31

In3C = Informantin 3 C-vaihe:

las2-12tt01te08lo36

**Lähteet**

- Denisyuk, Anna 2010. Sanavalintavirheiden tyypit ja taustaa suomenoppijoilla. *Julkaisematon pro gradu -tutkielma*. Turun yliopisto.
- Ellis, Rod 1985. *Understanding Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, Rod 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, Rod, Gary Barkhuizen 2005. *Analysing Learner Language*. Oxford: Oxford University Press.
- Goldschneider, Jennifer M., Robert M. DeKeyser 2005. Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. – *Language Learning* 55 (S1), 27–77. [doi:10.1111/j.0023-8333.2005.00295.x](https://doi.org/10.1111/j.0023-8333.2005.00295.x)
- Housen, Alex 2002. A corpus-based study of the L2-acquisition of the English verb system. – Sylviane Granger, Joseph Hung, Stephanie Petch-Tyson (Eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins, 77–118.
- Ivaska, Ilmari 2009. Eksistentiaalilauseen ilmiäiset edistyneiden suomenoppijoiden kirjoituksessa. *Julkaisematon pro gradu -tutkielma*. Turun yliopisto.
- Ivaska, Ilmari, Kirsti Siitonen 2009. Syntaktisesti koodattu oppijankielen korpus: mahdollisuuksia ja ongelmia. – Pille Eslon, Katre Õim (Toim.). *Korpusuuringute metodoloogia ja märgendamise probleemid*. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 11. Tallinn: Tallinna Ülikool, 54–71.
- Ivaska, Ilmari, Kirsti Siitonen 2010. Syntaktisesti annotoidun korpuksen soveltaminen lauserakenteen tarkasteluun. <http://www.uwasa.fi/midcom-serveattachmentguid-8f4acbc51affba94a86440ca1cfedf55/tyopajat%2E.pdf> (1.3.2011).
- Ivaska, Ilmari, Kirsti Siitonen (tulossa). Avainrakenneanalyysi: tapa tutkia oppijankielen lauserakennetta korpusvetoisesti. – Hyväksytty julkaistavaksi AFinLA-e-julkaisussa 2011.
- Jantunen, Jarmo H., Saana Piltonen 2009. Oppijansuomen ja viron sähköiset tutkimusaineistot. – *Virittäjä* 113, 449–458.
- Kemmer, Suzanne, Michael Barlow 2000. Introduction: A usage-based conception of language. – Michael Barlow, Suzanne Kemmer (Eds.). *Usage-Based Models of Language*. Stanford: CSLI, vii–xxviii.
- Kurki, Tommi 2005. Yksilön ja ryhmän kielen reaaliaikainen muuttuminen. Kielenmuutosten seuraamisesta ja niiden tarkastelussa käytettävistä

- menetelmistä. Suomalaisen kirjallisuuden seuran toimituksia 1036. Helsinki: Suomalaisen kirjallisuuden seura.
- Martin, Maisa 2010. Korpukset toisen kielen oppimisen tutkimuksessa: sudenkuoppia ja kompastuskiviä. <http://www.uwasa.fi/midcom-serveattachment-guid-8f4acbc51affba94a86440ca1cfedf55/tyopajat%2E.pdf> (1.3.2011).
- Niiranen, Leena 2010. Tapaustutkimus kolmen suomenoppijan kompleksisista verbikonstruktioista. – Lähivördlusi. Lähivertailuja 20, 155–190. [doi:10.5128/LV20.05](https://doi.org/10.5128/LV20.05)
- Pajunen, Anneli 2001. Argumenttirakenne. Asiantilojen luokitus ja verbien käyttäytyminen suomen kielessä. Suomi 187. Helsinki: Suomalaisen kirjallisuuden seura.
- Pienemann, Manfred 2005. An introduction to processability theory. – Manfred Pienemann (Ed.). Cross-Linguistic Aspects of Processability Theory. Amsterdam: Benjamins, 1–61.
- Ringbom, Håkan 1993. Near-nativeness and the four language skills: Some concluding remarks. – Håkan Ringbom (Ed.). Near-native Proficiency in English. English Department Publications 2. Turku: Åbo Akademi, 295–303.
- Salmi, Niina 2010. Virkkeiden ja lauseiden piirteitä edistyneiden oppijoiden teksteissä. Julkaisematon pro gradu -tutkielma. Turun yliopisto.
- TS = Saukkonen, Pauli, Marjatta Haipus, Antero Niemikorpi, Helena Sulkala 1979. Suomen kielen taajuussanasto. Helsinki: WSOY.
- Tarnanen, Mirja 2002. Arvioija valokeilassa: suomi toisena kielenä -kirjoittamisen arviointia. Jyväskylä: Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto.
- Tognini-Bonelli, Elena 2001. Corpus Linguistics at Work. Amsterdam: Benjamins.

**Kirsti Siitonen**

Turun yliopisto  
suomen kieli, kieli- ja käännöstieteiden laitos  
FI 20014 Turku, Finland  
[kisiito@utu.fi](mailto:kisiito@utu.fi)

**Jenny Niemelä**

Turun yliopisto  
suomen kieli, kieli- ja käännöstieteiden laitos  
FI 20014 Turku, Finland  
[niemela.jenny@gmail.com](mailto:niemela.jenny@gmail.com)

## What does a longitudinal study reveal of advanced Finnish learners' language proficiency?

KIRSTI SIITONEN, JENNY NIEMELÄ

University of Turku

Finno-Ugric languages are not widely used by non-native speakers; therefore, native speakers are not accustomed to non-native speakers mastering their language well, and it is not expected of them. It would however be important for the native communities of Finno-Ugric languages to respect their own mother-tongue and to be aware of the social and political significance and value of the fact that a number of non-native speakers are interested in reaching a high academic level in their language.

At the University of Turku there is under preparation an encoded corpus of Finnish advanced learners' texts by means of language technology. Such a corpus could be used for broad research into the variation and changes within learners' language proficiency as a whole, as opposed to merely changes in a certain aspect of their language proficiency. This corpus would also use texts of native speakers as comparison material. The Finnish learners in question are advanced students in the masters' program for Finnish and Finno-Ugric languages.

In this article we will make a suggestion on how to longitudinally study the language proficiency of advanced Finnish learners' writing. In such a study individuals' actual development can be examined. The study is comparable to variation studies and real-time methods. At the beginning of a two years observation period, the language skills of the students are already quite good; therefore, their development is not rapid. Most likely their language will not be much better at the final stage compared to the initial phase, but it will probably be different. This article resembles a case study with only three informants. The observation period comprises 7 months for informants 1, 2 and 3. From informant 3 we also have material from the subsequent 7 months; thus this material comprises 14 months after the beginning of observation.

Examples of our research questions: In comparison to native speakers' Finnish, how does the learners' Finnish differ in overall structure? (Cf. verb

frequencies, errors or tongue slips in morphology, errors in predicate constructions, and complex verb constructions.) Examples of our hypotheses: the selection of various verbs will extend, the frequency of the verb *olla* ('to be') will decrease. Language will be enriched structurally due to the accumulation of complex nominal forms. The decisive questions are: How rapidly does language proficiency develop? What are the main characteristics of this development?

About the results of this case study: The article will clarify how certain indicators behave during the time of the study, how they are interconnected, and what the indicators can reveal of language skill. An advanced learner's language skill, as understood in light of the original presuppositions of this article, may vary widely during the period of observation. It is possible that an unambiguous tendency of change cannot be observed. These apparent fluctuations in skill can be due to the nature of the texts produced as well as due to lexical requirements. In such cases variations in the texts may be caused by the same factors as among native speakers, rather than by limitations of language proficiency.

All informants have a rich vocabulary already at the outset of research, and all of them employ quite many complex nominal forms. The result of quantitative analysis is that informants' problems connected to predicates decrease during the seven months observation period. However, they do not employ infinite structures in their language as frequently as native speakers do. Errors in morphology and writing instead appear to be connected with the personal properties of a given speaker, such as carefulness, and their number does not diminish even while language proficiency improves in other respects. The period of observation is admittedly quite short. It is stated in the holistic qualitative analysis that at the end of the period of observation, the texts of all informants contain very intelligible language and correct patterns are to be found among all of them.

The results of our inquiry report various texts by individual writers and indicate how the texts of learners may deviate from those of native speakers. We also receive indications concerning what approaches or questions could be suitable for research from large corpora.

**Keywords:** academic learner; examination texts; indicators of language skills' development; near native speaker; change tendencies; Finnish as a second language; verb skills; corpus research; orthographic errors; tongue slips