



ILMARI IVASKA ja KIRSTI SIITONEN

## ***tehdessä*-konstruktio edistyneessä oppijansuomessa**

### **Korpusanalyysin ja oikeakielisyyсарviointien ristivalotus**

#### *1. Johdanto*

Monia suomen infinitiivikonstruktioita pidetään kielenoppimisen kannalta kompleksisina rakenteina. Tällaisten moniosaisien ja siksi morfologisesti mutkikkaiden rakenteiden muodostus vaatii samanaikaisesti sekä verbien että nominien taivutustyyppien tuntemista ja tunnistamista. Tämän lisäksi useat infinitiivikonstruktioita ovat abstrakteja ja kognitiivisesti kompleksisia sekä syntaktiselta funktioltaan että semanttiselta tulkinnaltaan: monet niistä nominalisoivat verbikonstruktioita ja kokonaisia sivulauseita sekä ilmaisevat aika- tai kausaalisuhteita. Tämä infinitiivikonstruktioiden oletettu monimutkaisuus heijastuu myös suomi toisena tai vieraana kielenä -opetuksen (S<sub>2</sub>) oppikirjoihin niin, että infinitiivikonstruktioita käsitellään oppikirjoissa vasta melko myöhäisessä vaiheessa, kun verbi- ja nominitaivutuksen päämekanismit ja keskeiset taivutustyyppit on käsitelty. Infinitiivikonstruktioiden opetustraditio sisältääkin ajatuksen kielitaidon kumulatiivisesta kehittymisestä siten, että infinitiivikonstruktioiden hallinta edistyy vähittäin, kun konstruktioiden osatekijät ovat oppijan hallinnassa.

Seuraamme tässä tutkimuksessa ennen kaikkea Siitosen (2014) avaamaa tutkimuslinjaa eli tarkastelemme inessiivisijaisen E-infinitiivikonstruktion esiintymiä S<sub>2</sub>-kirjoittajien ja ensikielenään suomea (S<sub>1</sub>) kirjoittavien akateemisissa teksteissä ja vertaamme näitä toisiinsa. Tarkastelemme konstruktion aktiivi- ja passiivivarianttia, jotta voimme havaita käytössä mahdollisesti esiin nousevat konstruktion eri merkitysten käyttöerot eli sen käyttöalan supistumat, laajentumat ja näiden väliset yhteydet. Pyrimme tarkentamaan fokusta siihen, missä määrin S<sub>2</sub>- ja S<sub>1</sub>-kirjoittajien välillä havaittavat käyttöerot ovat tekstilajikohtaisia ja missä määrin erot ovat yhteydessä konstruktion eri merkityksiin. Aineistomme on osa Edistyneiden suomenoppijoiden korpusta (LAS<sub>2</sub>; Ivaska 2014c). Toisaalta selvitämme, miten S<sub>1</sub>-puhujat kokevat konstruktion laadullisen kohdekielenmukaisuuden. Tätä varten olemme toteuttaneet kaksi kyselytutkimusta. LAS<sub>2</sub>:sta poimitut esimerkit havainnollistavat aktiivista (1) ja passiivista (2) konstruktioita; tarkastelunalaiset konstruktioita on lihavoitu (ks. myös VISK § 120).

- (1) Jos oppijan L2-kielen taito ei opetuksen *alkaessa* ole kovin hyvä (las2-vtto1vertto29)<sup>1</sup>
- (2) Lausesemantiikkaa *tarkasteltaessa* sanat jaetaan sisältö- ja muotosanoihin. (las2-vtto1vertto25)

Vertaamme toisiinsa myös toisiaan lähellä olevia akateemisen kirjoittamisen tekstilajeja ja S<sub>2</sub>- ja S<sub>1</sub>- kirjoittajien tekstejä. Pyrimme näin ottamaan huomioon havainnon, että akateemiseen kielenkäyttöön kuuluvien, keskenään samankaltaistenkin tekstilajien ominaispiirteissä saattaa olla suuria eroja (ks. esim. Biber 2006). Tutkimus nivoutuukin osaksi konteksti- ja tekstilajikohtaista näkemystä kielenkäytöstä ja -oppimisesta. Tämä puolestaan vastaa monilta osin käyttö pohjaista ajatusta kielijärjestelmästä.

Oppijankielen kompleksisuutta voidaan tarkastella myös sen kohdekielenmukaisuuden kautta. Näkemysmme mukaan kohdekielenmukaisuus on jatkumomaista ja kielen sisäinen vaihtelu osa itse kielijärjestelmää. Ottaaksemme tämän kielen sisäisen vaihtelun huomioon perustamme tulkintamme oman intuitiomme sijaan oikeakielisyy sarviointeihin pohjauttaen kielenkäyttäjien jaettuun näkemukseen. Kielenkäyttäjäkunnan todentamat poikkeamat kohdekielestä osoittavat laadullista horjuntaa vielä edistyneenkin kielitaidon tasoilla. Laajemman arvioijajoukon tulkinnat antavat monipuolisemman ja selvemmin todelliseen kielenkäyttöön perustuvan kuvan kuin tutkijoiden intuiotensa avulla tekemät tulkinnat.

Tutkimuskysymyksemme ovat seuraavat:

- 1) Minkälaista on inessiivisijaisten E-infinitiivikonstruktioiden käyttö edistyneiden suomenoppijoiden ja ensikielisten suomenpuhujien teksteissä? Käyttöä tutkitaan sekä määrällisesti että laadullisesti akateemisen kirjoittamisen eri tekstilajeissa.
- 2) Havaitsevatko ensikieliset suomenpuhujat edistyneiden suomenoppijoiden tuottamien E-infinitiivikonstruktioiden poikkeamia kohdekielestä ja ovatko heidän havaintonsa ja tulkintansa keskenään samankaltaisia?

Artikkeli rakentuu seuraavasti: luvussa 2 käsittelemme tutkimuksen teoreettista viitekehystä ja esittelemme lyhyesti aiempia tuloksia. Luvussa 3 kuvaamme tutkimuksen aineiston ja menetelmät. Luvussa 4 puramme korpus pohjaisen analyysin ja oikeakielisyy sarviointien tulokset. Oikeakielisyy sarviointien yhteydessä analysoimme myös laadullisesti tapauksia, jotka nousevat arvioinneissa esiin kohdekielen vastaisina konstruktiivoina. Lopuksi luvussa 5 arvioimme tulosten merkitystä, käyttämämme menetelmää ja sen soveltuvuutta oppijankielen tutkimukseen. Pyrimme lisäksi liittämään tulokset osaksi keskustelua, joka koskee kielen kompleksisuutta ja erityisesti sen suhdetta kielenoppimiseen ja edistyneeseen kielitaitoon.

1 Merkintä kertoo esimerkin sijainnin korpuksessa, ja *vert* koodin osana kertoo, että kyse on S<sub>1</sub>-aineistosta.

## 2. Teoreettinen viitekehys ja aiempi tutkimus

Lähestymme tutkimuskohdettamme kognitiivisen konstruktioikieliopin näkökulmasta. Näemme kielijärjestelmän muodostuvan abstraktiudeltaan vaihtelevista konstruktioista, jotka puolestaan koostuvat muodon ja merkityksen tai funktion toistuvasta ja vakiintuneesta yhteydestä (kognitiivisesta konstruktioikieliopista tarkemmin, ks. Goldberg 1995; 2006). Tämä järjestelmä on jatkuvassa muutoksessa ja perustuu todelliseen kielenkäyttöön ja siihen pohjautuviin todennäköisiin kielellisiin toimintatapoihin (käyttöpohjaisesta ajattelusta tarkemmin, ks. Hopper 1987; Bybee–Thompson 2007; Larsen–Freeman–Cameron 2008; Cadierno–Eskildsen 2015).

Oppijankielen tutkimuksessa kompleksisuuden käsitettä on pohdittu paljon. Bultén ja Housenin (2012, 27) mukaan kompleksisuus pitäisi kuitenkin määritellä nykyistä tarkemmin. Yhtenäistä linjaa ei ole helppo löytää; jo kielikohtaiset erot ovat suuret. Morfologisesti rikkaassa kielessä kompleksisuus on eriluonteista kuin esimerkiksi kielessä, jossa sanajärjestys ilmentää semanttisia eroja. Tässä tutkimuksessa kyse on Bultéa ja Housenia seuraten rakenteelliseen kompleksisuuteen kuuluvasta formaalisesta ja funktionaaliseen syntaktiseen kompleksisuuteen kuuluu morfologinen ja funktionaaliseen syntaktinen kompleksisuus. Nämä molemmat kuuluvat lingvistiseen kompleksisuuteen ja siinä erityisesti kieliopilliseen kompleksisuuteen (ks. Bulté–Housen 2012, 23, 25, 27, erityisesti kuvio 1).

Morfologian näkökulmasta kompleksinen konstruktio koostuu esimerkiksi monista taivutuselementeistä – rakenteen moninaisuudessa voi olla variaatiota pääluokan, infiniittisten tunnusten, sijamuotojen ja liitteiden valinnassa. Syntaksissa otetaan huomioon lauseenjäsenten ja lauseiden keskinäiset suhteet. Infiniittisillä rakenteilla on rakenteen sisäiset syntaktiset suhteet osien kesken sekä rakenteen ulkopuoliset viittaussuhteet pääasiallisesti virkkeeseen, johon rakenne on upotettu. Suhteet voivat rakentua päällekkäisesti niin, että yksi ilmaus on huomioitava usein eri tavoin. Tällöin ilmaus ilmentää useampaa kompleksisuutta. (Bulté–Housen 2012, 27, 31, 35, erityisesti kuvio 2.) Infiniittiset rakenteet sovittautuvat syntaktisesti hallitsevaan lauseeseen, jossa konstruktio täyttää myös semanttisen tehtävän. Tämä tehtävä saattaa poiketa morfologisesta perusfunktioista. Näin sekä rakenteellinen että semanttinen variaatio lisäävät osaltaan kompleksisuutta.

Aiemmassa tutkimuksessa rakenteellista kompleksisuutta on lähestytty monin eri mittarein, usein määrällisen runsauden näkökulmasta – esimerkiksi tutkimalla lauseiden sanamääriä (mittausmenetelmistä ks. esim. Wolfe–Quintero ym. 1998). Kuten Martin kuitenkin huomauttaa, lauseiden ja virkkeiden pituuden ja suhteiden laskeminen ei ole mielekästä suomen kaltaisissa kielissä, joissa on kompakteja mutta hyvin ilmaisuvoimaisia infiniitiivikonstruktioita (Martin 2013). Tämä huomio pätee erinomaisesti myös tässä tutkittavaan E-infiniitiivikonstruktioon.

Inessiivisijainen E-infiniitiivikonstruktio on monimutkainen niin morfologialtaan kuin syntaktis-semanttiselta funktioltaankin. Muoto koostuu sanavartalosta, infiniiviin *-e*-tunnuksesta sekä inessiivin sijapäätteestä *-ssa*. Passiivimuotoisissa konstruktioissa

infinitiivin tunnusta edeltää passiivin tunnus ja aktiivimuotoisissa sijapäätettä voi seurata omistusliite. (VISK § 120.) Esimerkki 3 kuvaa aktiivista omistusliitteetöntä, esimerkki 4 passiivista ja esimerkki 5 aktiivista omistusliitteellisestä konstruktiosta. Konstruktiolla saat-  
taa myös olla – useimmiten genetiivisijainen – subjekti, kuten esimerkin 3 *keskustelun*.

- (3) Keskustelun *edetessä* puheenvuorot voivat pidentyä ja lomittua. (las2-vtto1vto89)
- (4) – – vaikka se on yksi keskeinen ero vesistöitä *verrattaessa* toisiinsa. (las2-vtto1vto82)
- (5) Jaollinen objekti ei *lisääntyessään* muuta olomuotoaan, esim maitoa maitoa maitoa. (las2-vtto1vto88)

Konstruktion syntaktinen funktio on tyypillisesti adverbiaalimääräite, joka ilmaisee konstruktion ilmaiseman teon ajankohtaa (VISK § 495), joka on ainakin osin päällekkäinen finiittisen rakenteen kanssa (VISK § 543, § 544, § 1122). Konstruktion parafraasit ovat usein *kun*-alkuisia sivulauseita, kuten 'kun keskustelu etenee, puheenvuorot voivat [...]' (3), '[...] on yksi keskeinen ero, kun vesistöjä verrataan toisiinsa' (4) tai '[...] ei muuta olomuotoaan, kun se lisääntyy' (5). Herlinin (1998, 183) mukaan kaikilla *kun*-sanon käytötapoilla näyttää olevan motivoitunut yhteys temporaalisuuteen, joka on myös E-infinitiivikonstruktion prototyyppinen merkitystulkinta. (*Kun*-lauseen merkitystulkintoista ks. esim. Herlin 1998; VISK § 1121 – § 1123, § 1128, § 1134, § 1137.) E-infinitiivikonstruktio hahmottuu konstruktion ja sen kehyslauseen välisen semanttisen suhteen merkitsimeksi (vrt. VISK § 1121). ISK:n (VISK § 545) mukaan konstruktion semanttinen tulkinta on joskus lähempänä kontrastiivista suhdetta (VISK § 1102), kausaalisuhdetta (VISK § 1128) tai kuvattun tilanteen ehdollisuutta (VISK § 1134) tai kahden tilanteen rinnasteisuutta (VISK § 545). ISK (mp.) mainitsee konstruktion yhdeksi käyttötavaksi rinnastuksen kaltaisen kahden tasaveroisen tilanteen välisen suhteen kuvaamisen. Tämä on kuitenkin ilmiönä tekstuaalis-funktionaalinen, mistä syystä sitä ei ole syytä käsitellä rinnakkain konstruktion merkitystä kuvaavien käyttötapojen kanssa.

Nykykielessä konstruktio on Herlinin ja Kotilaisen (2005, 284–286) mukaan jakautunut syntaktisesti kahtia lausemaisiiin käyttöympäristöihin tai *tarvittaessa*-ilmauksen kaltaisiin merkitykseltään ja muodoltaan kiteytyneisiin leksikaalistuneisiin käyttötilanteisiin. Herlinin ja Kotilaisen (mp.) mukaan on osin harhaanjohtavaa puhua eri merkityksistä tai merkityksen eriytyemisestä, sillä kyse on heidän mukaansa ennemmin merkityksen haa-  
listumisesta yleiseksi kahden prosessin välisen suhteen ilmaisimeksi.

Konstruktion käyttö eri merkityksissä on muuttunut osana kirjoitetun suomen konventioiden muotoutumista, ja konstruktion eri merkitysten tyypillisuus riippuu edelleen tekstilajista. Ensinnäkin konstruktio on yleistynyt kirjakielen kehityksen aikana huomattavasti: Herlinin ja Kotilaisen (2005, 261) mukaan sen määrä viisinkertaistui vuoden 1642 raamatunkäännöksestä Lönnrotin 1800-luvun teksteihin mennessä. Lisäksi konstruktio esiintyy tutkijoiden käyttämässä murreaineistossa puhtaasti ajan adverbiaalina, niin että sillä viitataan tilanteisiin, jotka tunnetaan puhetilanteessa joko yleisen maailmantiedon, tapahtuman toistuvuuden tai prosessin aiemman maininnan vuoksi (mt., 265–266).

Ikolan, Palomäen ja Koiton (1989, 386) mukaan konstruktio esiintyy kirjoitetussa kielessä noin kymmenen kertaa niin usein kuin puhekielessä (19 esiintymää / 1 000 sanetta vs. 2 esiintymää / 1 000 sanetta; temporaalirakenteen käytöstä ks. myös Lindén 1961; 1971). Vastaava trendi välittyy myös tarkasteltaessa Kielipankin Korp-käyttöliittymän<sup>2</sup> avulla kaikkien E-infinitiivin inessiivien esiintymiä Kansalliskirjaston lehtikokoelman 1990- ja 2000-luvun aineistoissa ja Lauseopin arkiston murrekorpuksessa: sanomalehtiaineistossa muotoa esiintyy 0,85 / 1 000 sanetta, kun taas murreaineistossa muotoa esiintyy 0,3 / 1 000 sanetta. Niin Eskola (2002, 137) kuin Herlin ja Kotilainenkin (2005, 288) toteavat konstruktion ei-temporaalisen käytön kuuluvan ennen kaikkea asiategistien kirjoituskäytänteisiin, mikä taas saattaa liittyä muiden kielten nimenomaan kirjoitettuun suomeen kohdistuneeseen vaikutukseen (Herlin–Kotilainen 2004).

Todellisesta variaatiosta poiketen E-infinitiivikonstruktion käyttöä koskevat hyvän yleiskielen ohjeistukset ovat konservatiivisia ja ne korostavat konstruktion temporaalista merkitystä. Esimerkiksi Osmo Ikola toteaa vuonna 2000 ilmestyneessä *Pienessä nykysuomen oppaassa* seuraavasti: ”Ellei temporaalista lauseenvastiketta voi muuttaa temporaaliseksi sivulauseeksi, on lauseenvastiketta käytetty väärin” (Ikola 2000, 170). Samoin *Uusi kielioapas* suosittelee, että ”[y]leensä ei temporaalista lauseenvastiketta [– –] pitäisi käyttää jos-lauseen sijasta” (Itkonen–Maamies 2007, 83). Vastaavasti Kirjoittajan ABC-kortti-verkkosivulla<sup>3</sup> todetaan seuraavasti: ”Temporaalinen lauseenvastike korvaa *kun*-lausetta ja ilmaisee siten aikaa. Niinpä sitä ei pidä käyttää esimerkiksi korvaamassa *jos*-, *vaikka*- tai *koska*-lausetta.” Kielenkäyttöä ohjeistavat lähdeoteokset ovat siis verrattain kaukana konstruktion todellisesta käytössä esiintyvistä variaatiosta.

Käyttöpohjaisen oppimiskäsityksen mukaan oppija havainnoi vastaantulevien kielinilmausten keskinäisiä samankaltaisuuksia ja erilaisuuksia niin muodon, merkityksen kuin käytönkin osalta. Hän hahmottaa havaintojensa avulla analogisesti kulloisenkin muodon ja merkityksen sekä käyttöyhteyden kokonaisuuden ja kytkee sen näin asteittain osaksi omaa kielijärjestelmäänsä (ks. esim. Tomasello 2003). Käyttöpohjaisen ajattelun mukaan omaksumista täytyy edeltää riittävä syötös ja tunnistaminen eli opittavan muodon käytön havaitseminen. Konstruktioiden merkityksen oppimisen on osoitettu alkavan tyypillisesti rakenteiden prototyyppisistä merkityksistä ja näille merkityksille tyypillisistä lekseemeistä ja laajentuvan eri käyttöyhteyksiin (ks. esim. Goldberg ym. 2004; Goldberg–Casenheiser 2008; Martin ym. 2010). Osa syötöksestä saattaa tulla luokkahuoneympäristössä tai oppikirjojen kautta, mutta koska E-infinitiivikonstruktio on tyypillinen erityisesti kirjoitetussa kielessä, sen tyypillisen käytön oppiminen edellyttää kykyä lukea ja ymmärtää suomenkielisiä tekstejä.

Konstruktioiden määrällisesti ja laadullisesti kohdekielinen käyttö edellyttää siis taitoa muodostaa rakenteet ja kykyä ymmärtää niiden syntaktis-semanttiset funktiot sekä

2 <https://korp.csc.fi>

3 [http://webcgi.oulu.fi/oykk/abc/kielenhuolto/lauseopillisia\\_ilmioita/lauseenvastikkeet/temporaalinen\\_lauseenvastike/](http://webcgi.oulu.fi/oykk/abc/kielenhuolto/lauseopillisia_ilmioita/lauseenvastikkeet/temporaalinen_lauseenvastike/)

tyypillinen konteksti- ja tekstilajikohtainen käyttö. Perinteisesti kielenilmausten kohdekielenmukaisuutta on lähestytty lähinnä tutkijan oman intuition avulla. Kohdekielimäisyys ei kuitenkaan ole yksiselitteistä, ja esimerkiksi Bybee ja Eddington (2006) osoittavat, että ensikieliset kielenkäyttäjät arvioivat usein esiintyvät ja niiden kanssa merkitykseltään samankaltaiset muodot oikeakielisemmiksi kuin harvemmin esiintyvät mutta laadullisesti samanlaiset muodot. Myös Dąbrowska (2008) osoittaa konstruktioiden prototyyppisten muotojen saavan yleensä selvästi korkeamman oikeakielisyyssarvion kuin saman konstruktion vähemmän prototyyppisten muotojen. Tutkittaessa E-infinitiivin kaltaista kompleksista konstruktioita S<sub>2</sub>-näkökulmasta onkin mielenkiintoista nähdä, miten ensikieliset suomenpuhujat arvioivat edistyneiden suomenoppijoiden tuotoksia. Keskenään yhdenmukaiset arviot tukevat nähdäksemme ajatusta konstruktion laadullisesta tyyppillisyydestä ja arvioiden vaihtelu taas indikoi yksilöllistä vaihtelua. Alhaiset oikeakielisyyssarviot ovat siis ensisijaisesti epäprototyyppisyyden ja kohdekielenvastaisuuden indikaattoreita. Täten ne todennäköisesti osoittavat omaksumisen osittaisuutta ja sitä, että oppijan kielijärjestelmä poikkeaa joiltakin osin ensikielisten järjestelmästä.

Inessiivisijaista E-infinitiivikonstruktioita on käsitelty S<sub>2</sub>-kontekstissa myös aiemmin. Tämän artikkelin kannalta keskeisin on Siitosen (2014) tutkimus, sillä sen kohteena ovat konstruktion passiivimuodot edistyneessä oppijansuomessa ja aineisto on osin samaa kuin tässä tutkimuksessa. Siitonen osoittaa, että S<sub>2</sub>- ja S<sub>1</sub>-aineistojen välillä havaittu ero ei selity yksittäisten tai tiettyjen harvojen lekseemien käyttöerolla vaan erolla nimenomaan konstruktioiden käytössä. S<sub>2</sub>-kirjoittajista vain alle kolmannes käyttää konstruktioita, S<sub>1</sub>-tekstien kirjoittajista kuitenkin lähes puolet. Lisäksi konstruktioita käyttävistä S<sub>2</sub>-kirjoittajista yli puolella ainakin osa konstruktioista on laadullisesti kohdekielenvastaisia.

S<sub>2</sub>-näkökulmasta konstruktioita ovat käsitelleet myös Eklund (1996) ja Tolvanen (2004). Eklundin (1996) S<sub>2</sub>-aineisto koostuu Helsingin yliopiston Suomen kielen ja kulttuurin sekä muiden edistyneille tarkoitettujen suomen kielen kurssien ja S<sub>1</sub>-aineisto tieteellisen kirjoittamisen kurssien teksteistä. Aineistojen välillä ei ole huomattavia määrällisiä käyttöeroja (mt., liite 2), mutta erityisesti konstruktion viittaussuhteet näyttävät aiheuttavan laadullisia ongelmia konstruktion tulkinnassa (mt. 83). Tolvasen (2004) aineisto puolestaan koostuu ylioppilasaineista, joissa hän toteaa infiniittisten konstruktioiden olevan melko harvinaisia niin S<sub>2</sub>- kuin S<sub>1</sub>-kirjoitelmissakin – kummassakin osaa-aineistossa noin kolmanneksessa aineista on joko yksi tai ei ainuttakaan Tolvasen lauseenvastikkeeksi luokittelemaa infiniittistä konstruktioita (mt. 33).

Käyttöpohjaiseen kielikäsitykseen kuuluu ajatus kielen funktionaalisesta luonteesta; monet kielelliset valinnat riippuvat niiden käyttöympäristöstä (ks. esim. Sinclair 1991, 112). Tekstilaji onkin keskeinen tekijä kielen oppimisessa. Swalesin (1990, 58) mukaan tekstilajit ovat kommunikaatiotilanteita, joiden osanottajat jakavat samat kommunikatiiviset tavoitteet ja jotka ovat samankaltaisia rakenteellisesti, tyyllisesti ja sisällöllisesti sekä myös vastaanottajien osalta. Berkenkotterin ja Huckinin (1995, ix) mukaan tekstilajitietoisuus koostuu kielenkäyttäjän kyvystä toimia kielellisesti tyyppillisellä tavalla samankaltaisina toistuvissa tilanteissa. Nämä kielelliset toimintamallit puolestaan kehkeytyvät

Hylandin (2004, 55–56) mukaan nimenomaan käyttöpohjaisesti syötöksen ja siitä tehtävien havaintojen ja yleistysten avulla. Tyypillinen kielellinen toiminta vaatii siis sekä rakenteiden laadullista hallintaa että kykyä toimia tilannekohtaisesti odotetulla tavalla, ja oppimisen kannalta laadulliset taidot edeltävätkin kontekstisidonnaista tyypillisyyttä (ks. esim. Olshtain–Blum–Kulka 1985; Bardovi–Harlig–Dörnyei 1998).

Olemme tässä tutkimuksessa päätyneet seuraamaan Biberin ja Conradin (2009, 2) määritelmiä, joiden mukaan tekstilaji koostuu tietynlaisista teksteistä, kun taas rekisteri kuvaa niitä kielellisiä piirteitä, joita tyypillisesti käytetään tietynlaisissa tilanteissa tietynlaisten kommunikatiivisten tavoitteiden saavuttamiseksi. Akateemiset tekstit mainitaan usein esimerkkinä tietystä rekisteristä, vaikka tosiasiaassa akateemisen kielenkäytön sisäinen vaihtelu on runsasta niin siinä esiin nousevien rekisterien kuin tekstilajienkin osalta. Esimerkiksi Biber ja Conrad kuvaavat akateemista kirjoittamista seuraavasti:

[Academic writing is] written language that has been carefully produced and edited, addressed to a large number of readers who are separated in time and space from the author, and with the primary communicative purpose of presenting information about some topic (Biber–Conrad 2009, 32).

Kuvaavaa on, että jo tämä luonteeltaan yleinen määritelmä rajaa ulkopuolelle monia akateemiselle kirjoittamiselle tyypillisiä tekstilajeja, kuten kurssityönä tuotettavat esseet ja tenttivastaukset. Mahdollista vaihtelua synnyttäviä tekijöitä on runsaasti: tekstien ensisijainen funktio, kommunikaation ensisijaiset osallistujat ja heidän asemansa akateemisessa yhteisössä, tekstien julkisuus, tekstien tuottamistilanne ja osallistujien mahdollisuudet valmistautua tekstin tuottamiseen, mahdollisuus käyttää apuvälineitä sekä se akateeminen ala, johon tekstit kuuluvat (ks. esim. Biber ym. 2002; 2004; Biber 2006). Biber (2006) osoittaa, että niin leksikaalisten kuin kieliopillistenkin piirteiden vaihtelu on runsasta sekä erilaisten tekstilajien että myös eri alojen välillä. Akateemisessa kontekstissa käytettyä, ensikielenään jotakin muuta kuin englantia puhuvien kirjoittajien englantia onkin alettu etenkin viimeisen kymmenen vuoden aikana tutkia nimenomaan korpustutkimuksen menetelmin sisäisen variaation (ks. esim. Callies 2013; Callies–Zaytseva 2013; Chen 2013; Schutz 2013) ja usein myös opetukseen soveltamisen näkökulmasta (ks. esim. Conrad 2005; 2011; O’Keeffe ym. 2007; Aijmer 2009; Meunier 2012). Paquot ja muut (2013) osoittavat, että vielä edistyneetkin kielenoppijat ilmaisevat oman näkökulmansa esimerkiksi akateemisissa teksteissä ensikielisiä kielenkäyttäjiä eksplisiittisemmin ja useammin tieteenalasta riippumatta.

### 3. *Materiaali ja menetelmät*

Tutkimus jakautuu niin materiaalien kuin menetelmienkin osalta kahteen osaan – korpuspohjaiseen määrällis-laadulliseen analyysiin ja sen pohjalta koostettuun kyselytutkimukseen todellisten tekstilauseiden oikeakielisyyssarvioinneista. Esittelemme tästä syystä aineiston ja menetelmät kahdessa osassa.

### 3.1. Korpuspohjainen analyysi

Tarkastelemme LAS2:n (Ivaska 2014c) avulla kolmea akateemisen kirjoittamisen tekstilajia: tenttivastauksia (tentit), pro gradu -käsikirjoituksia (gradut) ja esseitä (esseet). Tekstilajit ovat monilta osin samankaltaisia, mutta ne poikkeavat toisistaan niin tuottamistilanteen, ensisijaisen funktion kuin oletetun lukijakunnankin osalta. Tentit on kirjoitettu valvotussa tilanteessa käsin, kun taas gradut ja esseet on kirjoitettu valvomattomissa olosuhteissa niin, että kirjoittajan on myös ollut mahdollista konsultoida lähteitä ja saada apua. Tenttien ensisijaisena funktiona on osoittaa kirjoittajan tiedot ja taidot opiskelusta sisällöstä, gradujen tavoitteena on osoittaa kirjoittajan kyky akateemisen tutkimuksen tekoon, kun taas esseissä kirjoittaja pyrkii kuvaamaan omia ajatusprosessejaan ja tulkintamalleja sekä muodostamaan perusteltuja mielipiteitä opiskelusta sisällöstä. Siinä missä tentit ja esseet ovat luonteeltaan yksityisiä ja tarkoitettu ensisijaisesti opettajan luettaviksi, gradu-käsikirjoitusten lopputuotteena julkaistavat gradut ovat saatavilla myös muiden luettaviksi. Tekstilajien rinnakkainen tarkasteleminen S2- ja S1-aineistossa helpottaa havainnoitavan ilmiön ymmärtämistä. Samalla se auttaa tulkitsemaan konstruktioiden käyttöön vaikuttavia tekijöitä ja infinitiivikonstruktioiden ja tekstilajin suhdetta kielen kompleksisuuteen. Taulukko 1 kuvaa käyttämämme korpusaineiston koon.

LAS2 on annotoitu, xml-muotoon tallennettu tekstikorpus, jonka tekstit on rakenteistettu kappaleisiin, virkkeisiin ja lauseisiin. LAS2 on digitoitu alkuperäisistä teksteistä manuaalisesti tai muuntamalla tietokoneohjelman avulla, koodattu kielellisiin metatiedoin lemmojen, sanaluokkien, morfologisten muotojen ja syntaktisten funktioiden osalta. Koodaus on kohdistettu sanatasolle, joten yhdellä sanalla saattaa olla samanaikaisesti useita morfologisia koodeja. (Ks. tarkemmin Ivaska 2014c.)

Taulukko 1. Korpusaineiston koko ja jakauma osa-aineistoittain.

	S2-aineisto		S1-aineisto	
	Tekstien määrä	Saneiden määrä	Tekstien määrä	Saneiden määrä
<b>Tenttivastaukset (TENTTI)</b>	120	70 552	120	58 630
<b>Pro gradu -käsikirjoitukset (GRADU)</b>	16	29 435	16	39 492
<b>Esseet (ESSEE)</b>	32	16 375	32	22 807
<b>Yhteensä</b>	<b>168</b>	<b>116 362</b>	<b>168</b>	<b>120 929</b>

Korpusanalyysimme on otteeltaan korpuspohjainen. Olemme siis valinneet aiemman tutkimuksen ja omien havaintojemme perusteella tietyn kielen konstruktion – inessiivimuotoisen E-infinitiivikonstruktion – jonka tyypillistä käyttöä ja käytössä esiintyvää vaihtelua tarkastelemme korpusaineiston avulla. Menetelmä on sekä kuvaileva että vertaileva, sillä tarkastelemme käyttöä ylipäätään ja myös eri osa-aineistoissa. Lisäksi vertailemme osa-aineistoja keskenään. Tältä osin tutkimuksemme seuraa oppijankielen korpus tutkimuksessa yleistä kontrastiivista välikielianalyysia (engl. *Contrastive Interlanguage Analysis* – CIA; Granger 1996; 2015). Vertailu on monitasoista: vertaamme toisiinsa sekä S2- ja S1-aineistoja että eri tekstilajeja.



Lähtökohtanamme ovat morfologisesti määriteltyjen E-infinitiivikonstruktioiden käyttöfrekvenssit. Käyttöfrekvenssien ja niihin perustuvien tulkintojen lisäksi perehdyimme konstruktion laadullis-funktionaaliseen vaihteluun. Pyrimme näin vastaamaan CIA-tutkimusta kohtaan esitettyyn kritiikkiin siitä, että menetelmää soveltavat tutkimukset keskittyvät yksittäisten kielenpiirteiden frekvenssieroihin ja jättävät huomiotta niiden käyttöympäristön ja sen sisällään pitämät kielelliset ja kielenulkoiset muuttujat (ks. esim. Gries–Deshors 2014). Tutkimme ennen kaikkea sitä, miten konstruktiota käytetään eri tekstilajeissa, mitkä ovat sille tyypillisiä leksikaalisia representaatioita ja mitä funktioita se tyypillisesti ilmaisee. Käytännössä korpusanalyysi käsittää seuraavat vaiheet:

1. Kootaan kooltaan tasapainoinen otos tarkasteltavien tekstilajien tekstejä siten, että niitä on S<sub>2</sub>- ja S<sub>1</sub>-aineistossa kustakin tekstilajista yhtä paljon;
2. Haetaan kaikki inessiivisijaisen E-infinitiivikonstruktion esiintymät;
3. Tehdään tilastollinen vertailu konstruktion käyttöfrekvensseistä S<sub>2</sub>- ja S<sub>1</sub>-aineistojen sekä eri osa-aineistojen välillä;
4. Käydään läpi kaikki konstruktiesiintymät kontekstissaan sekä tulkitaan ja koodataan aineistoon niiden ensisijainen merkitys. Muuttujan mahdolliset arvot ovat aikasuhte, syysuhte ja ehtosuhte (VISK § 543–§ 545);
5. Tarkastellaan konstruktion eri merkitysten käyttöaloja, skemaattisuutta ja leksikaalista jakaumaa eri osa-aineistoissa.

Konstruktioiden merkitysten luokittelu on väistämättä tulkinnanvaraista, ja nyt käsillä olevan kaltaisen monimerkityksisen tai merkitykseltään haalistuneen konstruktion tapauksessa tämä on vielä ilmeisempää. Kausaalisuus ja konditionaalisuus ovat jo itsessään jatkumomaisia ilmiöitä. Noordman ja de Blijzer (2000) osoittavat kausaalisten ilmausten merkityksen ymmärtämisen riippuvan samanaikaisesti sisällön ja episteemisten kytkösten välisestä suhteesta sekä synn ja vaikutuksen (engl. *cause and effect*) välisestä lineaarisesta ja myös käsitteellisestä järjestyksestä. Nyt käsillä oleva konstruktio kuvaa prototyypillisesti aikasuhdetta – myös kaikissa aineistomme konstruktioissa temporaalinen tulkinta on mahdollinen, ja esimerkiksi kiteytyneissä *mennessä* ja *kuluessa* -tapauksissa se lienee ainoa tulkinta. Muissa tapauksissa olemme valinneet ensisijaiseksi tulkinnaksi kontekstin parhaiten tukeman vaihtoehdon. Käytimme tulkinnan apuna kutakin merkitystä vastaavaa parafrasaa seuraavasti: konditionaalisenä parafrasina oli 'sillä ehdolla, että', kausaalisenä parafrasina 'siitä syystä, että' ja ne tapaukset, joihin kumpikaan parafraasi ei sopinut, tulkitsimme ensisijaisesti temporaalisiksi. Epäselvimmät tapaukset olisi nähdäksemme voinut jättää käsittelemättä tai soveltaa luokittelussa epäselvät-kategoriaa tai pyrkii löytämään todennäköisimmät ratkaisut. Olemme päätyneet viimeiseen vaihtoehtoon keskustelemalla monitulkintaisimmista tapauksista yhdessä ja valinneet tulkinnan, joka tyydyttää kumpaakin kirjoittajaa.

Kaikki hakutulokset on tallennettu LAS2-hakutyökalujen avulla csv-muotoon. Tilastolliset analyysit on tehty R-ohjelmointiympäristössä (R Core Team 2015). Käyttöfrekvenssien tilastollisissa tarkastelussa olemme käyttäneet tekstikohtaisia suhteellisia

frekvenssejä eli esiintymien määrää kussakin tekstissä tuhatta sanetta kohti. Frekvensseissä havaittujen tilastollisten erojen testinä olemme käyttäneet Mann-Whitneyn U-testiä, joka on luonteeltaan epäparametrinen eikä näin ole herkkä esiintymien epäsymmetrisille jakaumille. Konstruktion skemaattisuutta tarkastelemme vertaamalla konstruktion esiintymäfrekvenssiä eli konstruktioiden absoluuttista määrää ja tyyppifrekvenssiä – tässä tapauksessa sitä, kuinka monta eri verbilekseemiä konstruktiossa esiintyy (ks. Bybee–Thompson 1997). Leksikaalista jakaumaa ja eri osa-aineistoja toisistaan erottavia lekseemejä tarkastelemme distinktiivisen kollekseemianalyysin avulla. Karkeistaen analyysissa verrataan jonkin konstruktion ja jonkin sen piirteen – kuten tietyn lekseemin – välisen yhteyden vahvuutta vertaamalla näitä kummankin kokonaisuuteen aineistossa. Menetelmä on kehitetty alun perin funktioltaan samankaltaisten konstruktioiden leksikaalisemanttisten preferenssien tarkasteluun (Gries–Stefanowitch 2004). Sitä on käytetty tämän tutkimuksen tapaan myös saman konstruktion leksikaalis-semanttisten käyttöerojen tarkasteluun eri osa-aineistoissa (ks. esim. Gries–Hilpert 2008; Ivaska 2015).

### 3.2. Kyselytutkimukset

Kyselytutkimuksemme seuraavat aiempien käyttö pohjaisten oikeakielisyyssarviointien tutkimusasetelmaa (ks. esim. Dąbrowska 2004; 2008; Bybee–Eddington 2006). Toteutimme kaksi osin päällekkäistä kyselytutkimusta: kyselytutkimus 1 muodostaa pääasiallisen kyselytutkimusaineistomme, kun taas kyselytutkimus 2 tarkentaa joitakin tutkimuksen kannalta relevantteja kysymyksiä. Informantit vastasivat kyselyihin kynällä ja paperilla, ja kerätyt vastaukset on digitoitu csv-muotoon. Oman tulkintamme mukaan osa kyselyn virkkeistä on täysin kohdekielen mukaisia, kun taas osassa virkkeistä on jotakin yllättävää nimenomaan tutkitussa infinitiivikonstruktiossa.

#### Kyselytutkimus 1

Kyselytutkimuksessa 1 vastaajia pyydettiin arvioimaan 44 suomenkielisen LAS2-korpuksen virkkeen oikeakielisyyttä viisiportaisella asteikolla niin, että vastaajan mielestä (1) on yksiselitteisesti epäkieliopillinen ja (5) yksiselitteisesti kieliopillinen. Virkkeistä kymmenessä esiintyi tutkimuskohteemme E-infinitiivikonstruktio ja loppuissa jokin muu infinitiivikonstruktio. Näin pyrimme välttämään ns. priming-ilmiötä eli vastaajien virittymistä tulkitsemaan yhtä tiettyä konstruktioita. Vastaajille ei kerrottu, mihin virkkeiden ominaisuuteen heidän tulisi kiinnittää huomiota. Arvioiden jälkeen vastaajia pyydettiin järjestämään em. kymmenen tutkittavaa virkettä alenevaan järjestykseen kieliopillisimmasta epäkieliopillisimpaan. Tarkastellut konstruktioit näkyvät esimerkeissä 6–15.

- (6) Verkko-opetusta *rakentaessa* kannattaa ensin määritellä sen tärkeimpiä perusteita eli oppimiskäsityksiä ja didaktiikkaa.
- (7) Piirteen tehtävää on ollut vaikea luokitella kielioppia *kirjoitettaessa*.

- (8) **Opettaessa** äännteitä oppijoiden täytyy muistaa, että suomi on vokaalirikas kieli.
- (9) Ammattikouluun tai yliopistoon **siirtyessä** oppilaan pitää edelleen saada tukea.
- (10) EU:lla on todellakin maailman isoin kääntämis- ja tulkkauspalvelu kaikkien jäsenvaltioiden virallisten kielten **ollessaan** myös työkieliä.
- (11) Lapsen toinen kieli on käytössä vain kotona ja vain tietyistä asioista **keskustellessa**.
- (12) **Ääntäessä** ilmavirta pääsee kurkunpäästä esteettä ulos.
- (13) Maija vastaa **kysyttäessä**, että hän asuisi kyllä mieluummin Ahvenanmaalla.
- (14) Komponenttianalyysia **tehdessä** valitaan ensin sanaryhmä tai sanapari.
- (15) Kantakielestä nykysuomeen **siirtyessä** monta konsonanttia on kadonnut.

Tarkastelemme virkkeiden saamien oikeakielisyyssarvioiden keskiarvoa ja sitä, miten yhdenmukaisia eri vastaajien arviot ovat ja missä määrin vastaajien esittämät virkkeiden kieliopillisuusjärjestykset poikkeavat toisistaan. Lisäksi tarkastelemme sitä, miten nämä vastaukset suhteutuvat konstruktion semanttisiin funktioihin eri osa-aineistoissa. Kyselyaineisto on kerätty Turun alueella kevään 2015 aikana 23:lta suomea ensikielenään puhuvulta informantilta. Vastaajat olivat akateemisesti koulutettuja ja heistä 19 joko opiskeli tai työskenteli kielen parissa.

## Kyselytutkimus 2

Kyselytutkimus 2 täydentää kyselytutkimusta 1. Vastaajia pyydettiin järjestämään samat kymmenen virkettä (6–15) alenevaan järjestykseen kieliopillisimmasta epäkieliopillisimpaan. Tämän jälkeen vastaajille esitettiin samat virkkeet sekä aktiivi- että passiivimuodossa ja heitä pyydettiin arvioimaan virkkeiden oikeakielisyyttä viisiportaisella asteikolla niin, että (1) on yksiselitteisesti epäkieliopillinen ja (5) yksiselitteisesti kieliopillinen. Tämän jälkeen vastaajille esitettiin kustakin kymmenestä virkkeestä kolme eri parafrasaa – ’sillä ehdolla, että’, ’siitä syystä, että’ ja ’silloin, kun’ – ja heitä pyydettiin arvioimaan nämä parafrasit viisiportaisella asteikolla niin, että (1) on yksiselitteisesti mahdoton tai järjetön ja (5) taas kaikin puolin mahdollinen ja järkevä. Lopuksi vastaajia pyydettiin kommentoimaan alkuperäisiä virkkeitä (6–15) sen kannalta, mikä niissä on kohdekielenvastaista. Kyselyaineisto on kerätty Turun alueella syksyn 2016 aikana, ja siihen vastasi 35 informanttia, joista 25 puhuu ensikielenään suomea ja 10 jotakin muuta kieltä. Kaikki vastaajat olivat akateemisesti koulutettuja, ja he opiskelivat kieliaineita. Ensikielenään jotakin muuta kuin suomea puhuvia vastaajia voidaan pitää korpusaineistomme S2-informanttien tavoin edistyneinä kielenoppijoina, sillä he suorittivat opintojaan suomen kielellä.

#### 4. Tulokset

##### 4.1. Inessiivisijaisen E-infinitiivikonstruktion käytön määrällinen yleiskuva

Tutkimus- ja vertailuaineistossa on kaikkiaan 355 inessiivisijaista E-infinitiivikonstruktiota, joista 242 (0,98 / 1 000 sanetta) on aktiivissa ja 113 (0,52 / 1 000 sanetta) passiivissa. Konstruktioista 129 on S2-aineistossa ja 226 S1-aineistossa. Taulukko 2 havainnollistaa esiintymämääriä sekä kuvaa S2- ja S1-aineistojen välisten erojen tilastollista merkitsevyyttä sekä tekstilajeittain että yhtenä kokonaisuutena. Keskiarvot, mediaanit ja erojen tilastolliset merkitsevyydet on laskettu tekstikohtaisista esiintymämääristä siten, että kunkin tekstin esiintymät on normalisoitu tuhatta sanetta kohti.

Taulukko 2. Inessiivisijaisten infinitiivikonstruktioiden esiintyminen eri osa-aineistoissa.

	PÄÄ-LUOKKA	TENTTI		GRADU		ESSEE		YHTEENSÄ	
		S1 n=142	S2 n=39	S1 n=43	S2 n=60	S1 n=41	S2 n=30	S1 n=226	S2 n=129
	aktiivi	n=81	n=32	n=29	n=40	n=32	n=28	n=142	n=100
	passiivi	n=61	n=7	n=14	n=20	n=9	n=2	n=84	n=29
MED. / 1000	aktiivi	0	0	0,20	0,39	0	0	0	0
	passiivi	0	0	0,10	0,32	0	0	0	0
KA. / 1000	aktiivi	1,13	0,58	0,96	0,87	1,44	0,41	1,17	0,79
	passiivi	0,93	0,17	0,62	0,86	0,55	0,03	0,83	0,21
U-testi	aktiivi	p = .00153		p = .5984		p = .1245		p = .00129	
	passiivi	p < .001		p = .516		p = .02788		p < .001	

Kuten esiintymisfrekvenssien tunnusluvut osoittavat, konstruktio on verraten harvainen – ja kuten mediaanit osoittavat, sitä esiintyy yli puolessa teksteistä vain graduissa. Aktiivi on passiivia yleisempi kaikissa osa-aineistoissa niin S2- kuin S1-aineistoissakin. Konstruktioita käytetään tasaisimmin graduissa, kun taas muissa osa-aineistoissa vaihtelua on paljon. Erityisesti S2-aineistossa passiivikonstruktiota esiintyy lähes yksinomaan graduissa.

Aktiivi- ja passiivivariantti ovat kumpikin harvinaisempia S2-aineistossa kuin S1-aineistossa, ja ero on molemmissa tapauksissa tilastollisesti merkitsevä<sup>4</sup>. S2- ja S1-aineistojen välinen frekvenssiero on passiivikonstruktioiden osalta tilastollisesti merkitsevä tenteissä ja esseissä ja aktiivikonstruktioiden osalta ainoastaan tenteissä. Graduissa konstruktioita käytetään itse asiassa S2-aineistossa S1-aineistoa enemmän, joskaan ero ei ole tilastollisesti merkitsevä. Konstruktioiden käyttö ei siis erota yleisesti S1-kirjoittajia ja suomen taidoiltaan edistyneitä S2-kirjoittajia toisistaan, vaan ero perustuu tämän tutkimuksen aineiston valossa tekstilajiin – vaikka käsillä olevat tekstilajit ovatkin monin tavoin samankaltaisia.

4 Tilastollisen merkitsevyyden tulkinta on tässä artikkelissa tehty tasolla  $p < .05$ .

#### 4.2. Funktionaalinen ja leksikaalinen vaihtelu

Aineiston 355 konstruktioesiintymästä noin 46 % ilmaisee ensisijaisesti aikasuhdetta, noin 31 % ehtosuhdetta ja noin 21 % kausaalista vaikutussuhdetta. Loput esiintymät ovat merkitykseltään epäselviä. Koska epäselviä tapauksia on hyvin vähän, ne on jätetty määrällisen tarkastelun ulkopuolelle. Esimerkki 16 kuvaa esiintymää, jonka olemme tulkinneet ensisijaisesti aikasuhdetta kuvaavaksi, 17 kuvaa ehtosuhdetta ja 18 kausaalista vaikutussuhdetta.

- (16) Tenttivastauksia voidaan pitää tutkimuksen kannalta havainnollisena materiaalina, koska ne on tuotettu kielen kannalta spontaanissa tilanteessa ilman apuvälineiden käytön mahdollisuuksia (vrt. Granger 2002, 8–9). Aineistoa on yhdeksältä informantilta ja sitä on yhteensä 2575 virkeriviä. Informantit ovat syntyneet ja asuneet osan elämästään muualla kuin Suomessa. Aineistoa *kerättäessä kaikki informantit opiskelivat suomea Turun yliopistossa*. Informanttien ensikielet ovat englanti, komi, liettua, ruotsi, tšekki, venäjä ja viro. (las2-vtto2vertoo2)
- (17) Myös passiivilla ilmaistaan intentionaalisuutta: ”Meille naurettiin”, ”Vene kaadettiin”. Passiivi on aina elollistarkoitteinen eli passiivin ”rajaama” epämääräinen joukko koostuu elollisista olennoista. *Elollisista olennoista puhuttaessa on syytä pohtia tekojen intentionaalisuutta* - lauseessa ”Vene kaadettiin” kaatajan roolissa olevat henkilöt tai henkilö ovat harkinnan tuloksena päättänyt kaataa veneen. (las2-vtto1vert127)
- (18) Luonnehtimisen kannalta on myös ollennaista onko kieli ylhäällä tai alhaalla. Diagrammassa näkee, että i, y ja ö ovat suppeita vokaaleja, *eli kielenkärki niitä ääntäessä on ylhaalla*. e, ö ja o ovat puolisupeita ja ä, a laveita. (las2-14tto1teo1000)

Funktiot jakautuvat aktiivi- ja passiivikonstruktioissa hyvin eri tavoin, sillä aktiivimuotoisista noin 58 % (n = 142) ilmaisee aikasuhdetta, 12 % (n = 28) ehtosuhdetta ja noin 28 % (n = 71) kausaalisuhdetta, kun taas passiivimuotoisista noin 19 % (n = 22) ilmaisee

Taulukko 3. Inessiivisijaisten E-infinitiivikonstruktioiden funktiot osa-aineistoittain.

	PÄÄ-LUOKKA	TENTTI		GRADU		ESSEE		YHTEENSÄ	
		S1	S2	S1	S2	S1	S2	S1	S2
AIKA	aktiivi	35	20	17	30	18	20	70	70
	passiivi	11	2	5	2	2	0	18	4
KAUSAALI	aktiivi	26	7	8	8	11	8	45	23
	passiivi	5	1	0	1	1	0	6	2
EHTO	aktiivi	16	5	4	1	2	0	22	6
	passiivi	45	4	9	17	6	2	60	23

aikasuhdetta, noin 73 % ( $n = 83$ ) ehtosuhdetta ja noin 7 % ( $n = 8$ ) kausaalisuhdetta. Taulukko 3 kuvaa näiden funktioiden esiintymistä osa-aineistoittain.

Aktiivikonstruktio on passiivivarianttia yleisempi kahden teon ajallista suhdetta tai tekojen keskinäistä vaikutussuhdetta ilmaisemassa, kun taas passiivi on aktiivisia yleisempi tekojen välistä ehtosuhdetta ilmaisemassa. Aktiivin tilastollisesti merkitsevä käyttöero tenteissä toistuu kaikissa konstruktion funktioissa niin, että ero on suhteellisesti suurin kausaali- ja ehtosuhteissa. Kun verrataan ylipäätään S2- ja S1-aineistoa toisiinsa, nähdään, että konstruktion ensisijainen funktio aikasuhteen ilmaisijana on selvästi yleisempi S2-aineistossa (57 % kaikista esiintymistä) kuin S1-aineistossa (39 % kaikista esiintymistä). Passiivien tilastollisesti merkitsevä käyttöero tenteissä heijastuu niin ikään kaikissa konstruktion funktioissa, joskin se on selvin konstruktion yleisimmässä funktiossa – ehtosuhteen ilmaisijana. Passiivikonstruktion osalta S2- ja S1-aineistojen välinen käyttöero näyttää siis yhtäältä olevan tekstilajikohtainen mutta toisaalta myös merkityskohtainen – passiivia käytetään ilmaisemaan nimenomaan ehtosuhdetta, mikä puolestaan on tyypillistä tenteille.

Aktiivi- ja passiivikonstruktioiden välisen suhteen kannalta on mielenkiintoista, että S2-aineistossa kunkin funktion yleisempi muoto kattaa myös suuremman osan kaikista esiintymistä. Näin on sekä kussakin tekstilajikohtaisessa osa-aineistossa että koko aineistossa. Esimerkiksi kaikista aikasuhdetta ilmaisevista konstruktioista suurempi osa on aktiivisia ja S2-aineistossa osuus on suhteellisesti suurempi (71 aktiivisia ja 4 passiivisia) kuin S1-aineistossa (71 aktiivisia ja 18 passiivisia). Vastaavasti gradujen ehtosuhdetta ilmaisevista konstruktioista valtaosa on passiivisia, ja ero on S2-aineistossa suurempi (17 passiivisia ja 1 aktiivi) kuin S1-aineistossa (9 passiivisia ja 4 aktiivisia). Ainoa poikkeus ovat ehtosuhdetta kuvaavat konstruktiot tenteissä, joissa myös kokonaismäärä on S2-aineistossa hyvin pieni. Konstruktioiden prototyyppinen käyttö on siis S2-aineistossa yleisempää ja tekstilajikohtainen funktionaalinen vaihtelu vähäisempää. Tenteissä havaittava passiivisten ehtoa ilmaisevien konstruktioiden poikkeava jakauma ei täysin sovi tulkintaan, mutta poikkeus saattaa johtua siitä, että ehtoa ilmaiseva funktio on ylipäätään vasta kehkeytymässä osaksi S2-kirjoittajien spontaania kielirepertoaria.

Tämän aineiston valossa on mahdotonta sanoa, minkälainen näiden kahden muuttujan vaikutussuhde on luonteeltaan tai suunnaltaan. Onko kyseessä siis edistyneiden kielinoppijoiden tekstilajikohtaisen kielenkäytön oppimiseen liittyvä ilmiö, joka näkyy tekstilajille tyypillisen funktion käytössä, vai konstruktion tietyn funktion oppimiseen liittyvä ilmiö, joka näkyy sille tyypillisessä tekstilajissa? Selvää silti on, että yhden tekijän huomioon ottaminen ilman toista ei tee oikeutta tarkasteltavalle ilmiölle. Tekstien tuottamistilanteen kannalta on kiintoisaa, että erot heijastuvat juuri tenteissä, joka on tarkastelluista tekstilajeista kaikkein spontaanein, vähiten suunniteltu ja yksityisin. Funktioiden kannalta on puolestaan kiintoisaa, että erot ovat selvempiä konstruktion niissä funktioissa, jotka eivät vastaa yksinomaan sen ensisijaista funktiota ajallisten suhteiden ilmaisijana.

Konstruktion leksikaalinen vaihtelu osoittaa, että konstruktio on selvästi produktiivinen niin S2- kuin S1-aineistossakin. S2-aineiston 129 esiintymässä on 57 eri verbiä ja

S1-aineiston 226 esiintymässä 95 eri verbiä. Toisin sanoen S2-aineistossa konstruktion absoluuttisen esiintymäfrekvenssin ja sen leksikaalista vaihtelua kuvaavan tyyppifrekvenssin suhde on  $57/129 \approx 0.44$ , kun taas S1-aineistossa vastaava suhde on  $95/226 \approx 0.42$ . Konstruktion suhteellisessa skemaattisuudessa ei toisin sanoen ole eroa osa-aineistojen välillä.

Konstruktion yleisimpien verbien vertailu S2- ja S1-aineistossa osoittaa, että joukossa on yksiselitteisesti aikasuhdetta kuvaavia verbejä, kuten *kulua*, *mennä* ja *kehittyä*, aika- ja ehtosuhdetta ilmaisemaan käytettyjä kommunikaatioverbejä, kuten *puhua*, *arvioida* ja *lukea*, ja generisiä, yksinomaan kontekstin perusteella tulkintansa saavia verbejä, kuten *olla*. Yleisimpien lekseemien määrä näkyy taulukosta 4.

*Taulukko 4.* Inessiivisijaisten E-infinitiivikonstruktioiden yleisimmät lekseemit osa-aineistoittain. Taulukossa on kunkin osa-aineiston sellaiset kolme yleisintä lekseemiryhmää (rivit 1, 2 ja 3), joissa esiintymiä on enemmän kuin yksi.

	PÄÄ-LUOKKA			GRADU		ESSEE	
		S1	S2	S1	S2	S1	S2
1	aktiivi	4 (olla, puhua)	4 (kulua, ääntää)	4 (puhua)	5 (kulua, mennä)	5 (lukea)	6 (kulua)
	passiivi	14 (puhua)	3 (tutkia)	3 (arvioida)	4 (puhua)	3 (puhua)	2 (tarvita)
2	aktiivi	3 (alkaa, edetä, kehittyä, kulua, mennä)	3 (puhua)	4 (edetä, kulua, olla, tarkastella)	3 (kuunnella, käyttää, kääntää, taivuttaa)	2 (aloittaa, edetä, olla, tehdä, työskennellä)	3 (tehdä)
	passiivi	6 (arvioida)		2 (kerätä, käsitellä, opettaa, tehdä)	3 (käsitellä)		
3	aktiivi	2 (antaa arvioida, haluta, kommunikoida, kuvailla, oppia, tarkastella, tehdä, tulkata, tulla, vaatia)	2 (syntyä)	2 (pohtia, todeta)	2 (liittyä puhua, tutkia, vertailla)		2 (opettaa, opiskella)
	passiivi	5 (tutkia)			2 (käyttää, vertailla, tarvita)		

Leksikaalisesta vaihtelusta heijastuu sama tendenssi kuin funktionaalisesta vaihtelusta. Aikasuhdetta ilmaisevat verbit ovat suhteessa yleisempiä S2-aineistossa, kun taas S1-aineistossa korostuvat niin aika- kuin ehtosuhdettakin ilmaisemaan käytetyt kommunikaatioverbit, kuten *puhua*, *arvioida* ja *lukea*. Distinktiivisen kollekseemianalyysin avulla tehty S2- ja S1-aineiston leksikaalisen vaihtelun tarkastelu tukee tätä tulkintaa. Analyysin tulokset näkyvät taulukosta 5.

Taulukko 5. Ne inessiivisijaisen E-infinitiivikonstruktion verbit, jotka erottavat distinktiivisen kollekseemianalyysin perusteella parhaiten S2- ja S1-aineistoja toisistaan.

KOLLEKSEEMI	S1	S2	DISTINKTIIVISYYS
<i>kulua</i>	7 (14,01)	15 (7,99)	0.002263461
<i>arvioida</i>	12 (7,64)	0 (4,36)	0.004965944

Verbin *kulua* käyttö erottaa osa-aineistoja toisistaan vahvimmin niin, että se esiintyy S2-aineistossa likimain kaksi kertaa niin usein (15 esiintymää) kuin tasainen jakauma (n. 8 esiintymää) antaisi olettaa ja S1-aineistossa vastaavasti harvemmin (7 esiintymää vs. n. 14 esiintymää). Vastaavasti verbi *arvioida* esiintyy S1-aineistossa selvästi tasaista jakaumaa yleisemmin (12 esiintymää vs. n. 8 esiintymää) ja S2-aineistossa harvemmin (0 esiintymää vs. n. 4 esiintymää). Leksikaalisen vaihtelun tarkastelu siis tukee aiempaa tulkintaa siitä, että S2-aineiston E-infinitiivikonstruktioissa on vähemmän vaihtelua ja että konstruktion prototyyppiset merkitykset kattavat suuremman osan kaikista esiintymistä.

#### 4.3. Oikeakielisyyсарvioinnit kyselytutkimuksissa

E-infinitiivikonstruktioiden oikeakielisyyсарvioissa nousee esiin muutamia varteenotettavia seikkoja. Taulukossa 6 esimerkit on järjestetty hyväksyttävimmästä vähiten hyväksyttävään niiden järjestelytehtävässä saamien arvioiden keskiarvojen perusteella. Taulukko havainnollistaa siis oikeakielisyyсарvioiden keskiarvoja (kyselytutkimus 1) sekä järjestelytehtävässä saatujen vastausten keskiarvoja (kyselytutkimus 1 ja 2).

Tarkastellut kymmenen esimerkkiä jakautuvat arvioiden mukaan kolmeen ryhmään. Jakauma toistuu sekä esimerkikohtaisissa oikeakielisyyсарvioissa että S1-arvioijien listoissa esimerkkien suhteellisesta oikeakielisyydestä. Yksi esimerkki (taulukon 6 rivi 1) on luokiteltu selvästi muita oikeakielisemmäksi, ja se erottuu myös oikeakielisyyсарjestyksessä muista esimerkeistä. Toisen ryhmän muodostavat kuusi muuta esimerkkiä (taulukon 6 rivit 2–7), joiden arviot ovat keskenään samankaltaisia ja joiden keskinäinen järjestykseen perustuva luokitus tosin vaihtelee huomattavasti ja erottuu samalla muista esimerkeistä. Kolmannen ryhmän esimerkit (taulukon 6 rivit 8–10) on puolestaan arvioitu selvästi muita vähemmän oikeakielisiksi, ja niiden arviot ovat keskenään samankaltaisia.<sup>5</sup> Nämä kolme ryhmää hahmottuvat taulukon riviä 7 lukuun ottamatta samankaltaisiksi sekä kyselytutkimusten 1 ja 2 S1-aineistossa, kun taas kyselytutkimuksessa 2 kerätystä S2-aineistossa arvioitujen virkkeiden keskinäinen järjestyks on hyvin erilainen eikä se noudata edellä kuvattua ryhmittymistä.

Tulkintamme mukaan S1-arvioijat ovat verraten yksimielisiä siitä, että rivin 1 esimerkki on kohdekielenmukainen, rivien 2–7 esimerkeissä on jotain yllättävää mutta ne

<sup>5</sup> Kuten nämä arviot osoittavat, ensikielisetkään arvioijat eivät koskaan ole täysin yksimielisiä. Nyt käsitellyn konstruktion käytössä ja sen hallinnassa on vaihtelua ja horjuntaa myös ensikielisten kielenkäyttäjien piirissä (Kurki 2016).



Taulukko 6. Kymmenen E-infinitiivikonstruktion oikeakielisyyssarvioinnit.

	ESIMERKKI	KYSELYTUTKIMUS 1 (S1)		KYSELYTUTKIMUS 2 (S1 ja S2)	
		ARVION KA. (arviot 1 – 5)	JÄRJESTYKSEN KA. (järjestys 1 – 10)	S1  JÄRJESTYKSEN KA. (järjestys 1 – 10)	S2  JÄRJESTYKSEN KA. (järjestys 1 – 10)
1	<i>Piirteen tehtävää on ollut vaikea luokitella kieliooppia kirjoitettaessa.</i>	4.6	2.17=1	2.68=1	5.3=8
2	<i>Lapsen toinen kieli on käytössä vain kotona ja vain tietyistä asioista keskustellessa.</i>	3.9	3.52=2	4.44=5	4.6=6
3	<i>Komponenttianalyysejä tehdessä valitaan ensin sanaryhmä tai sanapari.</i>	4.0	3.70=3	4.28=4	3.5=2
4	<i>Ammattikouluun tai yliopistoon siirtyessä oppilaan pitää edelleen saada tukea.</i>	3.9	4.09=4	4.44=5	4.3=5
5	<i>Verkko-opetusta rakentaessa kannattaa ensin määritellä sen tärkeimpiä perusteita eli oppimiskäsityksiä ja didaktiikkaa.</i>	3.6	4.35=5	4.04=2	2.1=1
6	<i>Ääntäessä ilmavirta pääsee kurkunpäästä esteettä ulos.</i>	3.7	4.39=6	4.08=3	4.2=4
7	<i>Maija vastaa kysyttäessä, että hän asuisi kyllä mieluummin Ahvenanmaalla.</i>	3.9	4.91=7	5.48=7	7.2=10
8	<i>Kantakielestä nykysuomeen siirtyessä monta konsonanttia on kadonnut.</i>	2.9	6.26=8	5.96=8	4=3
9	<i>Opettaessa äänteitä oppijoiden täytyy muistaa, että suomi on vokaalirikas kieli.</i>	2.6	6.48=9	6.88=9	4.9=7
10	<i>EU:lla on todellakin maailman isoin kääntämis- ja tulkkauspalvelu kaikkien jäsenvaltioiden virallisten kielten ollessaan myös työkieliä.</i>	2.7	6.65=10	7.2=10	6=9

koetaan kuitenkin melko oikeakielisiksi, kun taas rivien 8–10 esimerkeissä on jotain selvästi kohdekielenvastaista. Kyselytutkimuksen 2 tulokset paljastavat lisäksi, että rivin 7 esimerkki on koettu kohdekielenvastaiseksi ennen kaikkea sanajärjestyksen ja *mieluimmin*-sanan vuoksi, ei infinitiivin takia. Virkettä 7 sekä S<sub>1</sub>- että S<sub>2</sub>-arvioijat pitivät selvästi kohdekielenvastaisena.<sup>6</sup> Muuten S<sub>2</sub>-arvioijien vastauksista samat kolme ryhmää eivät kuitenkaan nouse esiin eli S<sub>2</sub>-vastaajat eivät jaa samaa käsitystä virkkeiden kohdekielennukaisuudesta.

Ensikieliset arvioijat kokevat passiivikonstruktion aktiivikonstruktiota kohdekielennukaisemmaksi kaikissa tarkastelluissa virkkeissä taulukon 6 rivin 10 esimerkkiä lukuun ottamatta – passiivimuotojen arvioiden keskiarvo on kussakin tapauksessa viisiportaisella asteikolla yli 3.8. Aktiivimuodoista kohdekielennukaisimmaksi on arvioitu rivin 6 *Ääntäessä ilmavirta pääsee kurkunpäästä esteettä ulos*, ja sen saamien arvioiden keskiarvo on 3.01. Rivin 10 esimerkkivirke on todettu selvästi kohdekielennustaiseksi niin aktiivikuin passiivimuotoisenakin. Suomea toisena kielenä puhuvien vastauksista ei nouse esiin mitään selvää tendenssiä – puolessa lauseista (taulukon 6 rivit 2, 5, 6, 9, 10) aktiivi- ja passiivimuotoinen lause on koettu keskimäärin yhtä kohdekielennukaiseksi, kolmessa tapauksessa (taulukon rivit 3, 4, 8) aktiivimuoto on koettu selvästi passiivimuotoa kohdekielennukaisemmaksi ja kahdessa tapauksessa (rivit 1 ja 7) passiivimuoto on koettu aktiivimuotoa kohdekielennukaisemmaksi.

Kyselytutkimuksen 2 S<sub>1</sub>-vastaajat ovat verraten yksimielisiä siitä, mikä kussakin lauseessa on epäodotuksenmukaista tai kohdekielennustaiseista. Tarkastelluista kymmenestä virkkeestä kuudessa (taulukon 7 rivit 2, 3, 5, 6, 8 ja 9) ensikieliset arvioijat kokevat ongelmaksi E-infinitiivin pääluokan tai ainakin E-infinitiivin ylipäättään. Kahdessa tapauksessa (taulukon 7 rivit 4 ja 10) ongelman katsotaan liittyvän omistusliitteeseen ja kahdessa (taulukon 7 rivit 1 ja 7) ei valtaosan mielestä ole mitään kohdekielennustaiseista. Useassa tapauksessa iso osa S<sub>1</sub>-arvioijista ei ehdota yhtään korjausta huolimatta siitä, että arvioinneissa virkkeet arvioitiin jollakin tavalla kohdekielennustaisiksi. S<sub>2</sub>-arvioijat esittävät vielä vähemmän korjausehdotuksia; valtaosa heistä ei tarjoa minkäänlaisia korjauksia virkeisiin. Kun nämä arvioijat ehdottavat korjausta, se kohdistuu useimmiten infinitiiviin: he esimerkiksi ehdottavat sijalle jotakin toista infinitiivikonstruktiota, kuten MA-infinitiiviä. Omistusliitteen käyttöön liittyviä korjausehdotuksia ei ole kuin yksi kappale.

Sekä esimerkkikohtaiset että järjestykseen perustuvat arviot muodostavat samankaltaisen ryhmityksen, joten ensikieliset kielenkäyttäjät kokevat esimerkkilauseiden kieliopillisuuden keskenään samalla tavalla. Tarkastelemme seuraavaksi kolmea vähiten kohdekielennustaisena pidettyä esimerkkiä yksityiskohtaisemmin.

6 Kielenhuollonoppaissa käsitellään *essA*-infinitiivikonstruktiota melko vähän (ks. kuitenkin luku 2). Aktiivi- ja passiivikonstruktion hyväksyttävyyseroista olisi tärkeää tehdä ensikielisille kyselytutkimus vain ensikielisten tekstilauseista tai konstruoiduista tekstilauseista, joissa olisi sumea raja konstruktioiden käytön välillä. Oikeakielisyyttä ei aina voi ratkaista ilman tietoa kirjoittajan omista intentioista. Myös nollapersoonan ja ilmi-passiivin taustalla olevien tekijöiden tulkinta aiheuttaa sumeutta. Näitä kysymyksiä ei voi kuitenkaan pohtia enempää tässä artikkelissa.

Taulukko 7. Kymmenen E-infinitiivikonstruktion korjausehdotukset.

	ESIMERKKI	INFINIITIIVI+ PASSIIVI		INFINIITIIVI		INFINIITIIVI + OMISTUSLIITE		MUU		EI MITÄÄN	
		S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
1	<i>Piirteen tehtävää on ollut vaikea luokitella kielioppia kirjoitettaessa.</i>	1	2	1	1	0	0	2	1	21	6
2	<i>Lapsen toinen kieli on käytössä vain kotona ja vain tietyistä asioista keskustellessa.</i>	16	0	3	2	0	0	0	1	6	7
3	<i>Komponenttianalyysia tehdessä valitaan ensin sanaryhmä tai sanapari.</i>	12	3	1	1	0	0	0	0	12	6
4	<i>Ammattikouluun tai yliopistoon siirtyessä oppilaan pitää edelleen saada tukea.</i>	5/2	0	1	0	14/11	1	0	0	8	9
5	<i>Verkko-opetusta rakentaessa kannattaa ensin määritellä sen tärkeimpiä perusteita eli oppimiskäsityksiä ja didaktiikkaa.</i>	15	0	3	1	0	0	0	1	7	8
6	<i>Ääntäessä ilmavirta pääsee kurkunpäästä esteettä ulos.</i>	12	2	1	1	0	0	1	1	11	6
7	<i>Maija vastaa kysyttäessä, että hän asuisi kyllä mieluummin Ahvenanmaalla.</i>	1	1	0	0	0	0	5	2	19	7
8	<i>Kantakielestä nykysuomeen siirtyessä monta konsonanttia on kadonnut.</i>	17	2	4	2	0	0	1	1	3	5
9	<i>Opettaessa äänteitä oppijoiden täytyy muistaa, että suomi on vokaalirikas kieli.</i>	13	0	6	1	0	0	3	1	3	6
10	<i>EU:lla on todellakin maailman isoin kääntämisen ja tulkkauspalvelu kaikkien jäsenvaltioiden virallisten kielten ollessaan myös työkieliä.</i>	1	0	3	0	12	0	7	3	2	7

(19) Kantakielestä nykysuomeen **siirtyessä** monta konsonanttia on kadonnut. (las2-20tto1te01)

(20) **Opettaessa** äänteitä oppijoiden täytyy muistaa, että suomi on vokaalirikas kieli. (las2-21tto1te07)

Esimerkit 19 ja 20 (taulukon 6 rivit 8 ja 9) ovat S2-aineiston tenteistä. Niin oman arviomme kuin ensikielisten arvioijienkin mukaan esimerkkien kohdekielenvastaisuus liittyy ainakin pääosin konstruktion päaluokkaan – odotuksenmukainen muoto olisi passiivinen variantti. Konstruktiot voitaneen tulkita funktion kannalta kahdella tavalla. Aikasuhdetulkinta korostaisi konkreettista ajallista samanaikaisuutta tai lähekkäisyyttä

– esimerkissä 19 kehityksen etenemistä kantakielestä nykysuomeen ja esimerkissä 20 opettamisen ja muistamisen samanaikaisuutta. Toinen tulkintavaihtoehto on nähdä konstruktiot ehtosuhteen ilmaisijoina. Esimerkki 19 ilmaisisi sitä ehtoa, että tarkastelun fokus siirretään kantakielestä nykysuomeen. Esimerkissä 20 muistamisen ehtona olisi opetus-tilanne. Esimerkin 19 kohdalla kaikki arvioijat pitivät ehtosuhde- ja aikasuhte-parafraseja suunnilleen yhtä järkevinä (S1: ehtosuhde 3.3, kausaalisuhde 1.2 ja aikasuhte 3.6; S2: ehtosuhde 3.1, kausaalisuhde 1.7 ja aikasuhte 3.6<sup>7</sup>) ja esimerkin 20 kohdalla aikasuhte-parafrasi arvioitiin järkevimmäksi, joskin myös ehtosuhde koettiin täysin mahdolliseksi (S1: ehtosuhde: 3.6, kausaalisuhde 1.3 ja aikasuhte 4.6; S2: ehtosuhde 3.1, kausaalisuhde 2.0 ja aikasuhte 3.4). Ehtoa ilmaisevat konstruktiot lankeavat tenttivastausten osa-aineistossa ryhmään, jonka edustajia on S2-aineistossa vähän; viisi aktiivimuotoista ja neljä passiivimuotoista esiintymää. Ero S1-aineistoon on erityisen selvä passiivimuotoisten ehtoa ilmaisevien konstruktioiden osalta, sillä niitä on S1-aineiston tenttivastauksissa 45 esiintymää (aktiivimuotoisia on 16 esiintymää). Esimerkissä 20 on myös mahdollista, että kyseessä on *oppia*-verbin passiivin ja luonteeltaan kausatiivisen *opettaa*-verbin aktiivin yhteensulauma. Tulkinnan epäselvyys saattaa olla yksi keskeinen tekijä oikeakielisyys-arvioinnin kannalta ilmaistavan funktion ja sen tyypillisen muodon ristiriidan ohella.

- (21) EU:lla on todellakin maailman isoin kääntämis- ja tulkkauspalvelu kaikkien jäsenvaltioiden virallisten kielten *ollessaan* myös työkieliä. (las2-35tto1te02)

Esimerkin 21 konstruktion merkitys on kaukana aikasuhtetulkinnasta, ja se lähenee luonteeltaan kausaalisuhteen ilmaisemista niin, että esimerkin parafrasi olisi 'EU:lla on todellakin maailman isoin kääntämis- ja tulkkauspalvelu siitä syystä, että kaikkien jäsenvaltioiden viralliset kielet ovat myös työkieliä' (S1: ehtosuhde 2.0, kausaalisuhde 4.4, aikasuhte 2.9; S2: ehtosuhde 2.7, kausaalisuhde 4.1, aikasuhte 2.7). Esimerkin selkein kohdekielenvastainen piirre on possessiivisuffiksin käyttö tapauksessa, jossa viittauskohde ei ole ilmaistavissa persoonapronominilla (VISK § 1294). E-infinitiivikonstruktion ja kirjoitetun kielen substantiiveille ominaisen omistusliitteen käytön voineekin helposti nähdä oppimisen kannalta kompleksisena piirteenä, jossa ilmenee horjuntaa vielä edistyneessäkin oppijankielessä (Siitonen–Mizuno 2010). Toisaalta esimerkki 21 ilmaisee semanttisen kausaalisuhteen lisäksi myös eräänlaista tekstuaalista rinnastussuhdetta (VISK §545), mikä näyttää ainakin tämän aineiston valossa hyvin harvinaiselta. Tapaus voidaankin nähdä esimerkkinä samanaikaisesta useiden kielellisesti epätyypillisten piirteiden lankeamisesta yhteen, minkä on esitetty lisäävän kohdekielenvastaisten ilmausten mahdollisuutta (ks. Ivaska 2011). Esimerkistä 21 voidaan ajatella, että kompleksisen ja verraten harvinaisen konstruktion (-*essa*-konstruktio) epätyypillisen merkityksen (kausalisuhde) kanssa lankeaa yhteen kielenpiirre (possessiivisuffiksi), jonka tyypillinen käyttö rajautuu

7 Myös näissä arvioissa käytössä oli viisiportainen asteikko, jossa (5) vastaa täysin järkeenkäypää ilmausta ja (1) täysin järjetöntä ilmausta.

tiettyihin yhteyksiin (kirjoitettu kieli) ja jota käytetään tyypillisesti jonkin toisen sanaluokan (substantiivi) yhteydessä.

##### 5. Päätelmät, tulkinnat ja yhteenveto

Analyysin perusteella vastaamme tutkimuskysymyksiimme seuraavasti: Inessiivisijaisen E-infinitiivikonstruktion käyttö on kaikkiaan melko harvinaista – konstruktiota esiintyy alle puolessa tutkituista teksteistä. Konstruktion aktiivimuotoinen variantti on yleisempi niin S<sub>2</sub>- kuin S<sub>1</sub>-aineistossakin, joskin S<sub>2</sub>-aineistossa ero on S<sub>1</sub>-aineistoa suurempi. Tutkituista tekstilajeista konstruktion käyttö erottaa S<sub>2</sub>- ja S<sub>1</sub>-aineistoja toisistaan tilastollisesti merkitsevästi tenteissä sekä aktiivi- että passiivikonstruktioiden osalta ja esseissä passiivikonstruktioiden osalta. Kaikissa näissä tapauksissa konstruktio on yleisempi S<sub>1</sub>-aineistossa. Vajaa puolet konstruktioista ilmaisee ensisijaisesti aikasuhdetta, noin kolmannes ehtosuhdetta ja noin viidennes kausaalista vaikutussuhdetta. Nämä funktiot jakautuvat aktiivi- ja passiivikonstruktiioihin siten, että aikasuhte on selvästi yleisin merkitys aktiivikonstruktiolle, kun taas passiivikonstruktiosta valtaosa ilmaisee ehtosuhdetta. Tekstilajien välisistä eroista selvin on se, että ehtosuhteen ilmaiseminen on tyypillistä erityisesti tenteissä. S<sub>2</sub>- ja S<sub>1</sub>-aineistot eroavat toisistaan ennen kaikkea siten, että S<sub>2</sub>-aineistossa konstruktion tyypillisin merkitys – aktiivisissa tapauksissa aikasuhte ja passiivisissa tapauksissa ehtosuhte – kattaa suuremman osan konstruktion kaikista esiintymistä. Leksikaalinen variaatio tukee tätä tulkintaa, sillä ensisijaisesti aikasuhdetta ilmaisevat verbit ovat suhteellisesti yleisempiä S<sub>2</sub>-aineistossa, kun taas merkitykseltään yleisemmät verbit ovat suhteellisesti yleisempiä S<sub>1</sub>-aineistossa. Tarkasteltujen konstruktioiden merkitystulkinnat ovat väistämättä subjektiivisia ja riippuvat käytetyistä kriteereistä (tämän tutkimuksen kriteereistä ks. luku 3), kuten myös kyselytutkimuksemme tulokset osoittavat. Toisenlaisten kriteerien käyttö varmasti muuttaisi tuloksia tavalla tai toisella. Uskomme silti, että havaintomme funktion ja pääluokan korrelaatiosta samoin kuin tyypillisimpien funktioiden yleisyydestä ovat päteviä.

S<sub>1</sub>-puhujien oikeakielisyysarviot osoittavat, että he tunnistavat edistyneiden S<sub>2</sub>-oppijoiden tuottamien konstruktioiden poikkeamisen kohdekielestä ja kokevat oikeakielisyyden melko samalla tavalla. Kymmenestä tarkastellusta esimerkistä nousee esiin kohdekielisuuden kannalta kolme erilaista ryhmää: yksi esimerkki koettiin lähes yksiselitteisesti kohdekielenmukaiseksi, kuusi esimerkkiä vaikutti osin kohdekielenvastaisilta ja kolme esimerkkiä koettiin selvästi kohdekielenvastaisiksi. S<sub>2</sub>-arvioijien vastauksista nämä ryhmät eivät nouse esiin. S<sub>1</sub>-arvioijat olivat hyvin yksimielisiä siitä, että ongelmat liittyivät infinitiivin pääluokkaan tai joissakin tapauksissa omistusliitteen käyttöön, kun suomea toisena kielenä oppineiden arvioijien vastauksista tämä ei näy. On lisäksi merkillepantavaa, että kaikissa kolmessa kohdekielenvastaiseksi koetussa esimerkissä konstruktio ilmaisee jotakin muuta kuin sen ensisijaista, aikasuhdetta kuvaavaa merkitystä tai konstruktion merkitys on monitulkintainen. Emme ole tarkastelleet tulosten tilastollista merkitsevyyttä, sillä tällainen tarkastelu ei nähdäksemme merkittävästi vahvistaisi jo nyt

melko selviä havaittuja tendenssejä. S<sub>2</sub>-arvioijien vastauksia on sitä paitsi liian vähän tällaisen tarkastelun oikeuttamiseen.

Inessiivisijaisen E-infinitiivikonstruktion käytön kuvaamisen lisäksi tuloksilla on yleisempää merkitystä toisen kielen oppimisen tutkimukselle. Käytön suhteellinen vähäisyys ja laadullisesti kohdekielenvastaiset konstruktiot viittaavat osittaiseen omaksumiseen tai käynnissä olevaan omaksumisprosessiin. Kuten useat niin sanottua DEMfad-mallia (mallista ks. Martin ym. 2010) soveltavat tutkimukset osoittavat, kohdekielennukaista käyttöä edeltää lähes aina käytön frekvenssin kasvu (ks. esim. Kajander 2013; Seilonen 2013; Siivelt-Mustonen 2013). Prototyypisten merkitysten korostuminen S<sub>2</sub>-aineistossa vastaa aiemmassa tutkimuksessa (ks. esim. Lauranto 1997; Martin ym. 2010; Mustonen 2015) havaittua konkreettisten merkitysten vähittäistä vaihtelun kasvua ja abstraktistumista. Erot funktionaalissa distribuutiossa kuvannevat tällaista osittaista oppimista, sillä S<sub>2</sub>-aineistossa korostuvat S<sub>1</sub>-aineistoa enemmän konstruktion eri varianttien prototyyppisemmät muodot ja oikeakielisyyсарvioissa esiin nousseet kohdekielenvastaiset muodot ilmaisevat kaikki jotain muuta kuin konstruktion tyypillisintä merkitystä.

Metodologisesta näkökulmasta tutkimuksemme osoittaa, että oppijansuomen erityispiirteitä tarkasteltaessa on pyrittävä mahdollisuuksien mukaan huomioimaan oppimisprosessin dynaamisuus ja monitahoisuus (ks. esim. Larsen-Freeman 1997; Larsen-Freeman-Cameron 2008) ja prosessin monet muuttujat. Tässä tutkimuksessa tarkastelimme konstruktiota muotovariaation, leksikaalisen variaation, funktionaalisen variaation ja tekstilajien variaation näkökulmasta, joista kukin valottaa konstruktion käyttöä ja sen hallintaa osin eri tavalla ja joista yhdenkin poisjättäminen yksinkertaistaisi ja osin myös vääristäisi kokonaiskuva. Tämä on erityisen relevanttia kompleksisissa konstruktioissa, joiden muotoa ja merkitystä määrittävät samanaikaisesti useat mekanismit. Toinen metodologinen vahvuus on korpusmenetelmien yhdistäminen oikeakielisyyсарvioihin. Tieteellisessä keskustelussa on kaivattu korpustutkimuksen ja erilaisten kokeellisten menetelmien runsaampaa yhdistämistä (ks. esim. Gries 2012), ja näin tekemällä onkin saavutettu tuloksia, jotka olisivat jääneet saavuttamattomiin ilman monimenetelmäistä lähestymistä (ks. esim. Gries-Wulff 2005; 2009). Emme tämän tutkimuksen rajoissa paneudu kohdekielennukaaisuuden, kielellisen normin tai virheen käsitteisiin ja niiden määrittelyyn. Näemme kuitenkin, että käyttämämme menetelmä on luonteva ja käyttöpohjaiselle viitekehykselle perusteltu tapa ottaa huomioon nämä seikat ja silti välttää vertailuharha eli ajatus monoliittimaisesta staattisesta ja vaihteluttomasta kohdekielen järjestelmästä, johon oppijankieltä verrataan.

Konstruktion eri muotojen ja merkitysten käyttö on tekstilajikohtaista, sillä jopa tämän tutkimuksen aineistossa olleiden, keskenään verraten samankaltaisten tekstilajien välillä näyttää olevan selviä eroja konstruktion varianttien ja niiden funktioiden distribuutiossa. Käyttöpohjaisesta näkökulmasta on syytä kysyä, missä määrin konstruktion eri merkitysten oppiminen edellyttää niille tyypillisten käyttökontekstien tarvetta ja niille tyypillisten tekstilajien tuntemusta ja rekisterien hallintaa. Niin oppijanenglannin kuin -suomenkin osalta on osoitettu, että edistyneetkin oppijat käyttävät akateemisissa konteksteissa

tekstilajille epätyypillisiä kielenpiirteitä (Cotton 2010; Paquot ym. 2013; Ivaska 2014b). Kompleksisuuden kannalta voidaan vastaavasti kysyä, minkälainen on morfosyntaktisen kompleksisuuden ja kielen käyttökontekstien hallinnan välinen vaikutussuhde. Tulostemme valossa vaikuttaa siltä, että esimerkiksi E-infinitiivikonstruktion ehtoa ilmaiseva funktio on vasta kehkeytymässä osaksi S2-kirjoittajien spontaanissa tuottamistilanteessa – tässä tapauksessa tenteissä – käyttämää kielirepertoaria, vaikka sen käyttö S2-kirjoittajilla on muissa tarkastelluissa tekstilajeissa lähellä S1-kirjoittajien käyttöä. Tentit eroavat graduista ja esseistä niin suunnittelun ja huolitteluun kuin yksityisyyden ja odotettujen lukijoidenkin osalta. S2- ja S1-aineistojen välinen ero saattaa kuvata konstruktion käytön osittaisen oppimisen lisäksi myös tekstilajin ja sille tyypillisen rekisterin osittaista oppimista. Nähdäksemme konstruktiokohtaista oppimista tutkittaessa olisikin syytä ottaa aina huomioon ilmaistujen merkitysten monipuolistumisen (vrt. Seilonen 2013; Kajander 2013) lisäksi myös käyttökontekstien monipuolistuminen ja kullekin käyttökontekstille tyypillisen kielenkäytön tavat. Nämä näkökulmat kuvaavat saman ilmiön eri puolia – on kuitenkin käyttöpohjaisten mallien hengessä muistettava, että kaikki kielenkäyttö ja kielenkäytön tyypillisuus on kontekstikohtaista ja että konstruktioiden oppiminen on siis aina myös tekstilajien oppimista.

### Lähteet

- AIJMER, KARIN (toim.) 2009: *Corpora and Language Teaching*. Studies in Corpus Linguistics 33. John Benjamins, Amsterdam.
- BARDOVI-HARLIG, KATHLEEN – DÖRNYEI, ZOLTÁN 1998: Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *TESOL Quarterly* 32(2), 233–259.
- BERKENKOTTER, CAROL – HUCKIN, THOMAS N. 1995: *Genre knowledge in disciplinary communication*. Lawrence Erlbaum, Hillsdale (NJ).
- BIBER, DOUGLAS 2006: *University Language. A Corpus-Based Study of Spoken and Written Registers*. John Benjamins, Amsterdam. DOI: <http://dx.doi.org/10.1075/scl.23>.
- BIBER, DOUGLAS – CONRAD, SUSAN M. 2009: *Register, Genre, and Style*. Cambridge University Press, Cambridge. DOI: <http://dx.doi.org/10.1017/CBO9780511814358>.
- BIBER, DOUGLAS – CONRAD, SUSAN M. – REPPEN, RANDI – BYRD, PAT – HELT, MARIE 2002: Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly* 36(1), 9–48.
- BIBER, DOUGLAS. – CONRAD, SUSAN M. – REPPEN, RANDI – BYRD, PAT – HELT, MARIE – CLARK, VICTORIA – CORTES, VIVIANA – CSOMAY, ENIKO – URZUA, ALFREDO 2004: *Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus*. ETS TOEFL Monograph Series 25. Educational Testing Service, Princeton (NJ).
- BULTÉ, BRAM – HOUSEN, ALEX 2012: Defining and operationalizing L2 complexity. ALEX HOUSEN, FOLKERT KUIKEN ja INEKE VEDDER (toim.): *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA*, 21–46. John Benjamins, Amsterdam.
- BYBEE, JOAN – EDDINGTON, DAVID 2006: A usage-based approach to Spanish verbs of 'becoming'. *Language* 82(2), 323–355.
- BYBEE, JOAN – SANDRA THOMPSON 1997: Three frequency effects in syntax. *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Pragmatics and Grammatical Structure*, 65–85.
- CALLES, MARCUS 2013: Agentivity as a determinant of lexico-syntactic variation in L2 academic writ-

- ing. *International Journal of Corpus Linguistics* 18(3), 357–390. DOI: 10.1075/ijcl.18.3.05cal.
- CALLIES, MARCUS – ZAYTSEVA, EKATERINA 2013: The Corpus of Academic Learner English (CALE) – A new resource for the study and assessment of advanced language proficiency. SYLVIANE GRANGER, GAËTANELLE GILQUIN ja FANNY MEUNIER (toim.): *Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Corpora and Language in Use – Proceedings* 1, 49–59. Presses universitaires de Louvain, Louvain-la-Neuve.
- CADIERNO, TERESA – ESKILDSEN, SØREN WIND (toim.) 2015: *Usage-Based Perspectives on Second Language Learning*. Mouton de Gruyter, Berlin.
- CHEN, MEILIN 2013: A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics* 18(3), 418–442.
- CONRAD, SUSAN 2005: Corpus linguistics and L2 teaching. ELI HINKEL (toim.): *Handbook of research in second language teaching and learning*, 393–409. Lawrence Erlbaum, Mahwah, NJ.
- 2011: Variation in corpora and its pedagogical implications. VANDER VIANA, SONIA ZYNGIER ja GEOFF BARNBROOK (toim.): *Perspectives on corpus linguistics*, 47–62. John Benjamins, Amsterdam.
- COTTON, FIONA 2010: Critical thinking and evaluative language use in academic writing: A comparative cross-cultural study. GEORGE BLUE (toim.): *Developing Academic Literacy*, 73–85. Peter Lang, Oxford.
- DĄBROWSKA, EWA 2004: *Language, Mind and Brain. Some Psychological and Neurological Constraints on Theories of Grammar*. Edinburgh University Press, Edinburgh.
- 2008: Questions with long-distance dependencies: A usage-based perspective. *Cognitive Linguistics* 19(3), 391–425.
- EKLUND, MERJA 1996: *Temporaaliset sivulauseet ja lauseenvastikkeet ulkomaalaisopiskelijoiden kirjoitelmissa*. Sivulaudaturtutkielma. Helsingin yliopiston suomen kielen laitos.
- ESKOLA, SARI 2002: *Syntetisoivat rakenteet käännösuomessa. Suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmillä*. Joensuun yliopiston humanistisia julkaisuja 30.
- GOLDBERG, ADELE E. 1995: *Constructions. A Construction Grammar Approach to Argument Structure*. The University of Chicago Press, Chicago.
- 2006: *Constructions at Work – The Nature of Generalization in Language*. Oxford University Press, Oxford.
- GOLDBERG, ADELE E. – CASENHEISER, DEVIN 2008: Construction learning and second language acquisition. PETER ROBINSON ja NICK C. ELLIS (toim.): *Cognitive Linguistics and Second Language Acquisition*, 197–215. John Benjamins, Amsterdam.
- GOLDBERG, ADELE E. – CASENHEISER, DEVIN – SETHURAMAN, NITYA 2004: Learning argument structure generalizations. *Cognitive Linguistics* 15(3), 289–316.
- GRANGER, SYLVIANE 1996: From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. KARIN AIJMER, BENGT ALTENBERG ja MATS JOHANSSON (toim.): *Languages in Contrast*, 37–51. Lund University Press, Lund.
- 2015: Contrastive Interlanguage Analysis: A Reappraisal. *International Journal of Learner Corpus Research* 1(1), 7–24.
- GRIES, STEFAN TH. 2012: Corpus linguistics, theoretical linguistics, and cognitive/psycholinguistics: towards more and more fruitful exchanges. JOYBRATO MUKHERJEE ja MAGNUS HUBER (toim.): *Corpus linguistics and variation in English: Theory and description*, 41–63. Rodopi, Amsterdam.
- GRIES, STEFAN TH. – DESHORS, SANDRA C. 2014: Using regressions to explore deviations between corpus data and a standard/target: two suggestions. *Corpora* 9(1), 109–136.
- GRIES, STEFAN TH. – HILPERT, MARTIN 2008: The identification of stages in diachronic data: variability-based neighbor clustering. *Corpora* 3(1), 59–81.
- GRIES, STEFAN TH. – STEFANOWITSCH, ANATOL 2004: Extending collocation analysis: A corpus-based perspectives on 'alternations'. *International Journal of Corpus Linguistics* 9(1), 97–129.
- GRIES, STEFAN TH. – WULFE, STEFANIE 2005: Do foreign language learners also have constructions? Evidence from priming, sorting and corpora. FRANCISCO JOSÉ RUIZ DE MENDOZA IBÁÑEZ (toim.): *Annual Review of Cognitive Linguistics* 3, 182–200. John Benjamins, Amsterdam.



- GRIES, STEFAN TH. – WULFF, STEFANIE 2009: Psycholinguistic and corpus-linguistic evidence for L2 constructions. FRANCISCO JOSÉ RUIZ DE MENDOZA IBÁÑEZ (toim.): *Annual Review of Cognitive Linguistics* 7, 163–186. John Benjamins, Amsterdam.
- HERLIN, ILONA 1998: *Suomen kun.* Suomalaisen Kirjallisuuden Seura, Helsinki.
- HERLIN, ILONA – KOTILAINEN, LARI 2004: External factors behind cross-linguistic similarities. OLGA FISCHER, MURIEL NORDE ja HARRY PERRIDON (toim.): *Up and down the Cline. The Nature of Grammaticalization*, 263–279. Amsterdam: John Benjamins.
- 2005: Itsenäistyvä infinitiivi: 2. infinitiivin inessiivin kehitys kirjakielen aikana. ILONA HERLIN ja LAURA VISAPÄÄ (toim.): *Elävä kielioppi: suomen infinitiivien rakenteiden dynamiikkaa*, 258–291. Suomalaisen Kirjallisuuden Seura, Helsinki.
- HOPPER, PAUL 1987: Emergent grammar. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 139–157.
- HYLAND, KEN 2004: *Genre and Second Language Writing*. The University of Michigan Press, Ann Arbor (MI).
- IKOLA, OSMO 2000: *Pieni nykysuomen opas*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitos, Turku.
- IKOLA, OSMO – PALOMÄKI, ULLA – KOITTO, ANNA-KAISA 1989: *Suomen murteiden lauseoppia ja tekstikielioppi*. SKST 511. Suomalaisen Kirjallisuuden Seura, Helsinki.
- ITKONEN, TERHO – MAAMIES, SARI 2007: *Uusi kieliopas*. Tammi, Helsinki.
- IVASKA, ILMARI 2011: Lausetyyppien sekoittuminen edistyneessä oppijansuomessa – näkökulmana ek-sistentiaalilause. *Lähivertailuja. Lähivörtlusi* 21, 65–85. DOI: <http://dx.doi.org/10.5128/LV21.03>.
- 2014a: Edistyneen oppijansuomen avainrakenteita. Korpusnäkökulma kahden kielimuodon tyypillisiin rakenteellisiin eroihin. *Virittäjä* 118(2), 161–193. <http://ojs.tsv.fi/index.php/virittaja/article/view/9368/12373>.
- 2014b: Mahdollisuuden ilmaiseminen S1-suomea ja edistynyttä S2-suomea erottavana piirteenä. *Lähivertailuja. Lähivörtlusi* 24, 47–80. DOI: <http://dx.doi.org/10.5128/LV24.02>.
- 2014c: The corpus of advanced learner Finnish (LAS2): Database and toolkit to study academic learner Finnish. *Apples – Journal of Applied Language Studies* 8(3), 21–38.
- 2015: Longitudinal changes in academic learner Finnish: A key structure analysis. *International Journal of Learner Corpus Research* 1(2), 210–241. DOI: <http://dx.doi.org/10.1075/ijlcr.1.2.02iva>.
- KAJANDER, MIKKO 2013: *Suomen eksistentiaalilause toisen kielen oppimisen polulla*. Jyväskylän Studies in Humanities 220. Jyväskylän yliopisto, Jyväskylä.
- KURKI, TOMMI 2016: [Sähköpostiviesti]. Lähetetty 13.3.2016. Kirjoittajien hallussa.
- LARSEN-FREEMAN, DIANE 1997: Chaos/complexity and second language acquisition. *Applied Linguistics* 18(2), 141–165.
- LARSEN-FREEMAN, DIANE – CAMERON, LYNNE 2008: *Complex Systems and Applied Linguistics*. Oxford University Press, Oxford.
- LAURANTO, YRJÖ 1997: *Ensi askeleita paikallissijojen käyttöön. Espanjankielisten suomenoppijoiden sisä- ja ulkopaikallissijat konseptuaalisen semantiikan näkökulmasta*. Kakkoskieli 2. Helsingin yliopiston suomen kielen laitos.
- LINDÉN, EVA 1961: Temporaalirakenne suomen murteissa. *Virittäjä* 65(2), 194–210.
- 1971: Temporaalirakenne kirjoitetussa kielessä. *Virittäjä* 75(1), 23–34.
- MARTIN, MAISA 2013. Sentences and clauses as complexity measures in second language writing: a segmentation experiment. MARIA JÄRVENTAUSTA ja MARKO PANTERMÖLLER (toim.): *Finnische Sprache, Literatur und Kultur im deutschsprachigen Raum – Suomen kieli, kirjallisuus ja kulttuuri saksankielisellä alueella*, 185–198. VSUA 85. Harrassowitz, Wiesbaden.
- MARTIN, MAISA – MUSTONEN, SANNA – REIMAN, NINA – SEILONEN, MARJA 2010: On becoming an independent user. INGE BARTNING, MAISA MARTIN ja INEKE WEDDER (toim.): *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research*, 57–80. Eurosla: Monograph Series (1).
- MEUNIER, FANNY 2012: Formulaic language and language teaching. *Annual Review of Applied Linguistics* 32(1), 111–129.

- MUSTONEN, SANNA 2015: *Käytössä kehittyvä kieli: Paikat ja tilat suomi toisena kielenä -oppijoiden teksteissä*. Jyväskylä Studies in Humanities 255. Jyväskylän yliopisto, Jyväskylä.
- NOORDMAN, LEO – DE BLIJZER, FEMKE 2000: On the processing of clausal relations. ELIZABETH COUPER-KUHLEN ja BERND KORTMANN (toim.): *Cause - Condition - Concession - Contrast: Cognitive and Discourse Perspectives*, 35–56. Mouton de Gruyter, Berlin.
- O'KEEFE, ANNE – MCCARTHY, MICHAEL – CARTER, RONALD 2007: *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge University Press, Cambridge.
- OLSHTAIN, ELITE – BLUM-KULKA, SHOSHANA 1985: Degree of approximation: Nonnative reactions to native speech act behavior. SUSAN M. GASS ja CAROLYN G. MADDEN (toim.): *Input in second language acquisition*, 303–325. Newbury House, Rowley (MA).
- PAQUOT, MAGALI – HASSELGÄRD, HILDE – OKSEFJELL EBELING, SIGNE 2013: Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora. SYLVIANE GRANGER, GAËTANELLE GILQUIN ja FANNY MEUNIER (toim.): *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead. Proceedings of the First Learner Corpus Research Conference (LCR 2011)*, 377–387. Presses Universitaires de Louvain, Louvain-la-Neuve.
- R Core Team 2015: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.: <http://www.R-project.org/>.
- SCHUTZ, NATASSIA 2013: How specific is English for Academic Purposes? A look at verbs in business, linguistics and medical research articles. GISLE ANDERSEN ja KRISTIN BECH (toim.): *English Corpus Linguistics: Variation in Time, Space and Genre*, 237–257. Rodopi, Amsterdam.
- SEILONEN, MARJA 2013: *Epäsuora henkilöön viittaaminen oppijansuomessa*. Jyväskylä Studies in Humanities 197. Jyväskylän yliopisto, Jyväskylä.
- SIITONEN, KIRSTI 2014: Learners' dilemma: an example of complexity in academic Finnish. The frequency and use of the E infinitive passive in L2 and L1 Finnish. PEKKA LINTUNEN, MAIJA S. PELTOLA ja MARI-LIISA VARILA (toim.): *AFinLA-e Soveltavan kielitieteen tutkimuksia* 6, 134–148.
- SIITONEN, KIRSTI – MIZUNO, MANAMI 2010: Suomen monitahoinen possessiivisuffiksi ja suomenoppija. *Lähivertailuja. Lähivördlusi* 19, 136–159. DOI: <http://dx.doi.org/10.5128/LV19.09>.
- SIITONEN, KIRSTI – NIEMELÄ, JENNY 2011: Mitä pitkästä tutkimus paljastaa edistyneiden suomenoppijoiden kielitaidosta? *Lähivertailuja. Lähivördlusi* 21, 242–279. DOI: <http://dx.doi.org/10.5128/LV21.10>.
- SIIVELT, KEATY – MUSTONEN, SANNA 2013: Lähdekielen vaikutus ja kielitaitotasot: paikallissijojen kehitys oppijansuomessa. *Lähivertailuja. Lähivördlusi* 23, 341–370.
- SINCLAIR, JOHN 1991: *Corpus, Concordance, Collocation*. Oxford University Press, Oxford.
- SWALES, JOHN 1990: *Genre Analysis: English in Academic and Research Setting*. Cambridge University Press, Cambridge.
- TOLVANEN, MIIA 2004: *Lauseenvastikkeet ei-syntyperäisten kielenpuhujien ylioppilaisaineissa*. Pro gradu -tutkielma. Helsingin yliopiston suomen kielen laitos.
- TOMASELLO, MICHAEL 2003: *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Harvard University Press, Cambridge.
- VISK = AULI HAKULINEN – MARIA VILKUNA – RIITTA KORHONEN – VESA KOIVISTO – TARJA RIITTA HEINONEN – IRJA ALHO 2004: *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seura, Helsinki. Verkkoersio, <http://scripta.kotus.fi/visk> URN:ISBN:978-952-5446-35-7.

*ILMARI IVASKA – KIRSTI SIITONEN: The tehdessä construction in advanced learner Finnish*

Many Finnish infinitive constructions can be considered to be complex in terms of their morphological structure and their syntactic-semantic nature, and this has been thought to also affect the learning process of Finnish as a second language (F2) learners. This article focuses on the E infinitive construction (e.g. *tehdessä* and *tehtäessä*) and its use in the different genres of academic texts of F2 learners in comparison to those of native (F1) learners of Finnish. The approach is corpus-based, with data from the Corpus of Advanced Learner Finnish. Following the usage-based models of language, we supplement the corpus analysis with a grammaticality judgement survey taken by F1 and F2 learners. The research questions are as follows: 1) What is the use of E infinitive constructions like quantitatively and qualitatively in different genres of F2 and F1 academic writing? 2) Do F1 learners of Finnish pay attention to potential ungrammaticality of F2 texts and are their observations uniform?

The article shows that the construction is, in general, relatively rare and that the active construction is more frequent in both F2 and F1 data. As far as genres go, the construction is more rarely used in exam essays and research papers of F2 learners than F1 learners. Most commonly, the construction expresses a temporal relationship, although it is used also to express a conditional as well as a causal relationship. The temporal relationship is typical for the active constructions whereas the conditional relationship dominates the use of passive constructions. The most remarkable difference between the two genres is that the conditional relationship is especially typical for the exam essays. The most notable difference between the F2 and F1 data is that the most typical functions cover a greater part of all the occurrences in the F2 data than they do in the F1 data. The grammaticality judgements show that F1 learners recognize ungrammatical constructions relatively similarly, but that there are different degrees of ungrammaticality. In all of the constructions considered clearly ungrammatical, the function is not prototypical for the construction or it is unclear.

The results suggest that acquisition expands from more typical expressions to greater variation. Furthermore, the level of acquisition and the genre at hand inherently are intertwined. Less typical use may also be more prone to ungrammaticality. The method used shows that combining different methodologies gives a more detailed and fine-grained view on learning, and that leaving any of the now implemented points of view out of the analysis would falsely simplify a multi-faceted phenomenon.

Ilmari Ivaska

ilmari.ivaska@unibo.it

Dipartimento di Interpretazione e Traduzione - Università di Bologna

Corso della Repubblica 136

47121 Forlì, Italia

Kirsti Siitonen

kirsti.siitonen@utu.fi

Suomen kieli ja suomalais-ugrilainen kielentutkimus

20014 Turun yliopisto