

Summary: *tehdessä* construction in advanced learner Finnish

Many of the Finnish infinitive constructions can be considered complex in terms of their morphological structure and their syntactic-semantic nature, and this has been thought to affect also the learning process by 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 and the data come from the Corpus of Advanced Learner Finnish. Following the usage-based models of language, we complement the corpus analysis with a grammaticality judgement survey 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 relatively rare in general and that the active construction is more frequent both in F2 and F1 data. As far as genres go, the construction is more rarely used in the exam essays and in the research papers by F2 learners than by F1 learners. Most commonly, the construction expresses temporal relationship, although it is used also to express conditional as well as causal relationship. Temporal relationship is typical for the active constructions whereas conditional relationship dominates the use of passive constructions. The most remarkable difference between the genres is that conditional relationship is typical especially for the exam essays. The most notable difference between F2 and F1 data is that the most typical functions cover a greater part of all the occurrences in F2 data than they do in F1 data. The grammaticality judgements show that F1 learners recognize ungrammatical constructions relatively similarly but that there are different degrees of ungrammaticality. In all the constructions considered clearly ungrammatical, the function is not prototypical for the construction or it remains unclear. The results suggest that acquisition expands from the more typical expressions to greater variation. Furthermore, the level of acquisition and the genre at hand inherently intertwine. 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.

***tehdessä*-konstruktio edistyneessä oppijansuomessa:**

korpusanalyysin ja oikeakielisyyssarviointien ristivalotus

Ilmari Ivaska & Kirsti Siitonen

1. Johdanto

Monia suomen infinitiivikonstruktioita pidetään kielenoppimisen kannalta kompleksisina rakenteina. Ne ovat moniosaisia ja voidaan siten nähdä morfologisesti monimutkaisina, sillä muodostusmekanismit perustuvat lingvistisen analyysin kannalta samanaikaisesti sekä verbien että nominien taivutustyyppien tuntemiseen. Tämän lisäksi useat infinitiivikonstruktiot ovat abstrakteja ja kognitiivisesti kompleksisia sekä syntaktiselta funktioltaan että semanttiselta tulkinnaltaan; monet niistä nominalisoivat verbikonstruktioita

ja kokonaisia sivulauseita sekä ilmaisevat ajallisia tai kausaalisia suhteita. Tämä infinitiivikonstruktioiden oletettu monimutkaisuus heijastuu usein myös suomi toisena tai vieraana kielenä -opetuksen (S2) 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 kehitymisestä niin, että myös infinitiivikonstruktioiden hallinta edistyy vähittäin, kun konstruktioiden osatekijät ovat oppijan hallinnassa.

Tarkastelemme tässä artikkelissa kahta suomen infinitiivikonstruktiota – inessiivisijaisen E-infinitiivin aktiivista ja passiivista varianttia. Esimerkki (1) havainnollistaa aktiivista ja esimerkki (2) passiivista konstruktiota; tarkastelunalaisten konstruktiot on lihavoitu (ks. myös VISK § 120).

(1) *Jos oppijan L2-kielen taito ei opetuksen **alkaessa** ole kovin hyvä.* (las2-vtt01vert029)¹

(2) *Lausesemantiikkaa **tarkasteltaessa** sanat jaetaan sisältö- ja muotosanoihin.* (las2-vtt01vert025)

Nyt käsillä oleva tutkimus on osin jatkoa aiemmalle tutkimukselle (Siitonen 2014), jossa tarkasteltiin passiivimuotoisten E-infinitiivien käyttöä edistyneessä oppijansuomessa. Tässä tutkimuksessa laajennamme tarkastelun aktiivi- ja passiivimuotoisten konstruktioiden rinnakkaiseen tarkasteluun. Tämän lisäksi vertaamme toisiaan lähellä olevia akateemisen kirjoittamisen tekstilajeja ja S2-kirjoittajien ja ensikielenään suomea puhuvien (S1) kirjoittajien tekstejä toisiinsa. Pyrimme näin ottamaan huomioon havainnon siitä, että akateemisen kielenkäytön joukkoon kuuluvien, keskenään samankaltaistenkin tekstilajien ominaispiirteissä saattaa olla suuriakin eroja (ks. esim. Biber 2006). Liitämme näin havaintomme osaksi konteksti- ja tekstilajikohtaista näkemystä kielenkäytöstä ja -oppimisesta. Tämä puolestaan vastaa monilta osin käyttöpohjaista ajatusta kielijärjestelmästä. Toisaalta haluamme suhteuttaa havaintomme kielen kompleksisuuteen oppimisen kannalta tarkastelemalla käytettyjen konstruktioiden oikeakielisyyttä ensikielisten suomenpuhujien oikeakielisyyсарviointien valossa. Tutkimus nojaa ajatukseen kielijärjestelmän käyttöpohjaisesta perusluonteesta. Näkemyksemme mukaan kohdekielenmukaisuus on

¹ Merkintä kertoo esimerkin sijainnin korpuksessa.

jatkumomaista ja kielensisäinen vaihtelu sisäsyntyinen osa itse kielijärjestelmää. Ottaaksemme tämän kielensisäisen vaihtelun huomioon perustamme tulkintamme oman intuitiomme sijaan kielenkäyttäjien jaettuihin näkemyksiin. Kielenkäyttäjäkunnan validoimat poikkeamat kohdekielestä osoittavat laadullista horjuntaa vielä edistyneenkin kielitaidon tasoilla. Laajemman arvioijajoukon tulkinnat antavat monipuolisemman ja selvemmin todelliseen kielenkäyttöön pohjautuvan kuvan kuin tutkijoiden yksin 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?

Tämä artikkeli rakentuu seuraavasti: luvussa 2 käsittelemme tutkimuksen teoreettista viitekehystä ja esittelemme lyhyesti aiempia tuloksia. Luvussa 3 kuvaamme tutkimuksen aineiston ja menetelmät. Luvussa 4 puramme korpuspohjaisen analyysin ja oikeakielisyyssarviointien tulokset. Oikeakielisyyssarviointien yhteydessä analysoimme myös laadullisesti tapauksia, jotka nousevat arvioinneissa esiin kohdekielenvastaisina konstruktioina. Lopuksi luvussa 5 arvioimme tulosten merkitystä, käyttämämme menetelmän vahvuuksia ja soveltuvuutta oppijankielen tutkimukseen. Pyrimme lisäksi liittämään havaintomme osaksi keskustelua kielen kompleksisuudesta ja erityisesti sen suhteesta kielenoppimiseen ja edistyneeseen kielitaitoon.

2. Teoreettinen viitekehys ja aiempi tutkimus

Lähestymme tutkimuskohdettamme kognitiivisen konstruktiokieliopin näkökulmasta. Näemme kielijärjestelmän muodostuvan abstraktiudeltaan vaihtelevista konstruktioista, jotka puolestaan koostuvat muodon ja merkityksen tai funktion toistuvasta ja vakiintuneesta yhteydestä (kognitiivisesta konstruktiokieliopista tarkemmin, ks. Goldberg 1995; 2006). Uskomme tämän järjestelmän olevan luonteeltaan käyttöpohjainen siten, että kielijärjestelmä on jatkuvassa muutoksessa ja että se 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. 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 yhtäältä rakenteellisen kompleksisuuden piiriin kuuluvasta morfosyntaktisesta kompleksisuudesta ja toisaalta funktionaalista syntaktisesta ja fraseologisesta kompleksisuudesta. Nämä molemmat kuuluvat lingvistisen kompleksisuuden piiriin ja siinä erityisesti kieliopilliseen kompleksisuuteen (ks. Bulté–Housen 2012, 23, 25, 27 sekä erityisesti kuvio 1, s. 23).

Bulté ja Housen korostavat sitä, että kompleksisuus pitäisi määritellä tarkoin ja läpinäkyvästi. Ilmiö on monikasvoinen, ja sen eri osa-alueet tulisi pystyä määrittelemään yksiselitteisesti ja erottelemaan sekä ryhmittelemään. Morfosyntaktinen kompleksisuus sisältää morfologisen ja syntaktisen osan. Morfologinen kompleksisuus on teoreettisesti strukturaalisen kieliopillisen kompleksisuuden osa. (Bulté–Housen 2012, 27; kuvio 2.) Tietty konstruktio koostuu esimerkiksi monista taivutuselementeistä – rakenteen moninaisuudessa voi olla variaatiota pääluokan, infiniittisten tunnusten, sijamuotojen ja liitteiden valinnassa.

Kieliopillisen kompleksisuuden toisena teoreettisena osana on systeeminen kompleksisuus, jonka alakohtana on syntaktinen kompleksisuus. Infiniittisillä rakenteilla on rakenteen sisäiset syntaktiset suhteet eri osissa ja osien kesken sekä rakenteen ulkopuoliset viittaussuhteet pääasiallisesti virkkeeseen, johon rakenne on upotettu, mutta myös muuhun lähikontekstiin. Suhteet saattavat rakentua päällekkäisesti niin, että sama ilmiö joudutaan ottamaan huomioon eri lähtökohdista käsin ja se näin ollen ilmentää useampaa kompleksisuutta. (Bulté–Housen 2012, 27, 35.) Esimerkiksi tässä tutkimuksessa käsiteltävät syntaktisesti infiniittiset rakenteet sovittautuvat hallitsevaan lauseeseen, jossa konstruktio täyttää myös semanttisen tehtävän. Tämä tehtävä saattaa puolestaan poiketa morfosyntaktisesta perusfunktioista. Näin sekä funktionaalinen että semanttinen variaatio lisäävät osaltaan kompleksisuutta. Aiemmassa tutkimuksessa rakenteellista kompleksisuutta on lähestytty monin eri mittarein – usein siitä lähtökohdasta, että runsaus on merkki kompleksisuudesta (mittausmenetelmistä ks. esim.

Wolfe-Quintero ym. 1998). Kuten Martin kuitenkin huomauttaa, lauseiden ja virkkeiden pituuden ja suhteiden laskeminen ei ole mielekäästä suomen kaltaisissa kielissä, joissa on kompakteja mutta hyvin ilmaisuvoimaisia infinitiivikonstruktioita (Martin 2013).

Nähdäksemme tämä huomio pätee erinomaisesti myös nyt käsillä olevan tutkimuksen kohteeseen eli E-infinitiivikonstruktion.

Tämän tutkimuksen kohteena on siis inessiivisijainen E-infinitiivikonstruktio. Kuten todettua, konstruktio on verraten monimutkainen niin morfologialtaan kuin syntaktis-semanttiselta funktioltaankin. Muoto koostuu sanavartalosta, infinitiivin *-e*-tunnuksesta sekä inessiivin sijapäätteestä *-ssa*. Passiivimuotoisissa konstruktioissa infinitiivin tunnusta edeltää passiivin tunnus, ja aktiivimuotoisissa konstruktioissa sijapäätettä voi seurata omistusliite. (VISK § 120.) Esimerkki 3 kuvaa aktiivista omistusliitteetöntä konstruktiota, esimerkki 4 passiivista konstruktiota ja esimerkki 5 aktiivista konstruktiota, jossa on omistusliite. Konstruktioilla saattaa myös olla subjekti, joka on usein genetiivisijainen, kuten esimerkin 3 *keskustelun*.

(3) *Keskustelun edetessä puheenvuorot voivat pidentyä ja lomittua.* (las2-vtt01vert089)

(4) -- *vaikka se on yksi keskeinen ero vesistöitä verrattaessa toisiinsa.* (las2-vtt01vert082)

(5) *Jaollinen objekti ei lisääntyessään muuta olomuotoaan, esim maitoa maitoa maitoa.*
(las2-vtt01vert088)

Konstruktion syntaktinen funktio on tyypillisesti adverbiaalimäärite, joka ilmaisee ajankohtaa temporaalirakenteessa (VISK § 495) – tarkemmin sanottuna konstruktion ilmaiseman teon vähintään osittaista samanaikaisuutta finiittisen rakenteen kanssa (VISK § 543, § 544, § 1122). Konstruktion parafraasit ovat usein *kun*-alkuisia sivulauseita, kuten esimerkissä 3 'kun keskustelu etenee, puheenvuorot voivat [...]', esimerkissä 4 '[...] on yksi keskeinen ero, kun vesistöjä verrataan toisiinsa' tai esimerkissä 5 '[...] ei muuta olomuotoaan, kun se lisääntyy'. Kuten Herlin (1998, 183) toteaa, kaikilla *kun*-sanana käytötavoilla näyttää olevan motivoitunut yhteys temporaalisuuteen, joka on myös nyt käsillä olevan E-infinitiivikonstruktion prototyyppinen merkitystulkinta. (*kun*-lauseen merkitystulkinnosta ks. esim. Herlin 1998; VISK § 1121 – § 1123, § 1128, § 1134, § 1137.) Laajasti tarkastellen E-infinitiivikonstruktio hahmottuu konstruktion ja sen kehyslauseen välisen semanttisen suhteen merkitsimeksi (vrt. lauseiden välisten merkityssuhteiden tulkinnasta VISK § 1121). Ison suomen kieliopin (VISK § 545) mukaan konstruktion semanttinen tulkinta on joskus lähempänä kontrastiivista suhdetta

(VISK § 1102), kausaalisuhdetta (VISK § 1128) tai kuvatun tilanteen ehdollisuutta (VISK § 1134) tai kahden tilanteen rinnasteisuutta (VISK § 545). Iso suomen kielioppi (mp.) mainitsee konstruktion yhdeksi käyttötavaksi rinnastuksen kaltaisen kahden tasaveroisen tilanteen välisen suhteen kuvaamisen. Tämä on kuitenkin nähdäksemme 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 yhtäältä lausemaiseihin käyttöympäristöihin ja toisaalta merkitykseltään ja muodoltaan kiteytyneisiin, *tarvittaessa*-ilmauksen kaltaisiin leksikaalistuneisiin käyttötilanteisiin. Herlinin ja Kotilaisen (mp.) mukaan on osin harhaanjohtavaa puhua eri merkityksistä tai merkityksen eriytymisestä, sillä kyse on heidän mukaansa enemmän merkityksen haalistumisesta yleiseksi kahden prosessin välisen suhteen ilmaisimeksi.

Konstruktion käyttö sen eri merkityksissä on muuttunut osana kirjoitetun suomen konventioiden muotoutumista, ja sen eri merkitysten tyypillisuus riippuu edelleen kulloinkin käsillä olevasta tekstilajista. Ensinnäkin konstruktio on yleistynyt kirjakielen kehityksen aikana huomattavasti, sillä Herlinin ja Kotilaisen (2005, 261) mukaan sen määrä on viisinkertaistunut vuoden 1642 raamatunkäännöksen jälkeen Lönnrotin 1800-luvulla kirjoittamiin teksteihin mennessä. Lisäksi konstruktio esiintyy tutkimuksessa käytetyssä murreaineistossa puhtaasti ajan adverbialina niin, että sillä viitataan sellaisiin 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 likimain kymmenen kertaa niin usein kuin puhekielessä (19 esiintymää / 1000 sanetta vs. 2 esiintymää / 1000 sanetta; temporaalirakenteen käytöstä puhutussa ja kirjoitetussa kielessä, ks. myös Lindén 1961; 1971). Vastaava trendi välittyy myös verrattaessa Kielipankin Korp-käyttöliittymän² avulla kaikkien E-infinitiivin inessiivien esiintymiä Kansalliskirjaston lehtikokoelman 1990- ja 2000-luvun aineistoissa ja Lauseopin arkiston murrekorpuksessa, joskin jonkin verran heikompana: sanomalehtiaineistossa muotoa esiintyy 0,85 / 1000 sanetta, kun taas murreaineistossa muotoa esiintyy 0,3 / 1000 sanetta. Niin Eskola (2002, 137) kuin Herlin ja Kotilainenkin (2005,

² <https://korp.csc.fi>

288) toteavat konstruktion ei-temporaalisen käytön kuuluvan ennen kaikkea asiatekstien kirjoituskäytänteisiin, mikä saattaa puolestaan liittyä muiden kielten vaikutukseen nimenomaan kirjoitettuun suomeen (Herlin–Kotilainen 2004).

Käytön todellisesta variaatiosta huolimatta hyvän yleiskielen ohjeistukset ovat E-infinitiivikonstruktion käytön suhteen konservatiivisempia, ja niissä korostetaan konstruktion temporaalista merkitystä. Esimerkiksi Osmo Ikola toteaa vuonna 2000 ilmestyneessä Pienessä nykysuomen oppaassa seuraavasti: ”Ellei temporaalista lauseenvastiketta voi muutta temporaaliseksi sivulauseeksi, on lauseenvastiketta käytetty väärin” (Ikola 2000, 170). Samoin Uusi kieliopas suosittelee, että ”[y]leensä ei temporaalista lauseenvastiketta [-] pitäisi käyttää *jos*-lauseen sijasta” (Itkonen–Maamies 2007, 83). Vastaavasti Kirjoittajan ABC-kortti-verkkosivulla³ todetaan seuraavasti: ”Temporaalinen lauseenvastike korvaa kun-lausetta ja ilmaisee siten aikaa. Niinpä sitä ei pidä käyttää esimerkiksi korvaamassa *jos*-, *vaikka*- tai *koska*-lausetta.” Voitaneenkin sanoa, että kielenkäyttöä ohjeistavat lähdeteokset ovat verrattain kaukana konstruktion todellisesta käytössä esiintyvistä variaatiosta.

Kielen konstruktioiden oppiminen on nähdäksemme luonteeltaan käyttöpohjaista. Oppija siis hahmottaa vastaantulevien kielenilmausten keskinäisiä samankaltaisuuksia ja erilaisuuksia niin muodon, merkityksen kuin käytönkin osalta. Hän muodostaa havaintojensa avulla analogisesti kulloinkin kyseessä olevan muodon ja merkityksen sekä kulloisenkin 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 kulloinkin kyseessä olevan muodon käytön havaitseminen. Konstruktioiden merkityksen oppimisen on osoitettu alkavan tyypillisesti niiden prototyyppisistä merkityksistä ja näille merkityksille tyypillisistä lekseemeistä ja laajentuvan edelleen eri käyttöyhteyksiin (ks. esim. Goldberg ym. 2004; Goldberg–Casenheiser 2008; Martin ym. 2010). Oppimiskontekstin mukaan osa syötöksestä saattaa tulla luokkahuoneympäristössä tai oppikirjojen kautta, mutta – kuten edellä toteamme – E-infinitiivikonstruktio on tyypillinen erityisesti kirjoitetussa kielessä, mistä syystä sen

3

http://webcgi.oulu.fi/oykk/abc/kielenhuolto/lauseopillisia_ilmioita/lauseenvastikkeet/temporaalinen_lauseenvastike/

tyypillisen käytön oppimisen voidaan katsoa edellyttävän kykyä lukea ja ymmärtää suomenkielisiä tekstejä.

Konstruktioiden määrällisesti ja laadullisesti kohdekielinen käyttö edellyttää siis sekä taitoa muodostaa konstruktioita että kykyä ymmärtää niiden syntaktis-semanttiset funktiot ja tyypillinen konteksti- ja tekstilajikohtainen käyttö. Perinteisesti kielenilmausten kohdekielenmukaisuutta on lähestytty lähinnä tutkijan oman intuition avulla.

Kohdekielimäisyys ei kuitenkaan usein 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. Vastaavasti Dąbrowska (2008) osoittaa konstruktioiden prototyyppisten muotojen saavan yleensä selvästi korkeamman oikeakielisyyssarvion kuin saman konstruktion vähemmän prototyyppisten muotojen. Tutkittaessa E-infinitiivin kaltaista kompleksista konstruktioita S2-näkökulmasta onkin mielenkiintoista nähdä, miten ensikieliset suomenpuhujat arvioivat edistyneiden kielenoppijoiden tuotoksia. Nähdäksemme keskenään yhdenmukaiset arviot tukevat ajatusta konstruktion laadullisesta tyypillisyydestä tai epätyypillisyydestä ja vastaavasti arvioiden vaihtelu osoittaa, että kyseessä voi olla yksilöllinen vaihtelu. Näin ollen alhaiset oikeakielisyyssarviot ovat ensisijaisesti epäprototyyppisyyden ja kohdekielenvastaisuuden indikaattoreita. Täten ne voivat myös indikoida omaksumisen osittaisuutta ja sitä, että oppijan kielijärjestelmä poikkeaa joltakin osin ensikielisten kielenkäyttäjien järjestelmästä.

Inessiivisijaista E-infinitiivikonstruktioita on käsitelty S2-kontekstissa myös aiemmin. Tämän artikkelin kannalta keskeisin on Siitosen (2014) tutkimus, sillä sen kohteena on konstruktion passiivimuotojen käyttö edistyneessä oppijansuomessa ja aineisto on osin samaa kuin tässä tutkimuksessa. Siitosen mukaan konstruktioita käytetään S2-aineistossa huomattavasti S1-aineistoa harvemmin. Siitonen osoittaa myös, että havaittu ero ei selity millään yksittäisen tai tiettyjen harvojen lekseemien käyttöerolla vaan erolla nimenomaan konstruktioitasen käytössä. Omaksumisen osittaisuudesta kertonee se, että S2-kirjoittajista vain alle kolmannes käyttää konstruktioita, S1-tekstien kirjoittajista kuitenkin lähes puolet. Lisäksi on huomionarvoista, että konstruktioita käyttävistä S2-kirjoittajista yli puolella ainakin osa konstruktioista on laadullisesti kohdekielenvastaisia.

S2-näkökulmasta konstruktioita ovat käsitelleet myös Eklund (1996) ja Tolvanen (2004). Eklundin (1996) S2-aineisto koostuu Helsingin yliopiston Suomen kielen ja kulttuurin sekä muiden edistyneille tarkoitettujen suomen kielen kurssien teksteistä ja S1-aineisto tieteellisen kirjoittamisen kurssien teksteistä. Aineistojen välillä ei ole huomattavia määrällisiä käyttöeroja (mt., liite 2), mutta erityisesti konstruktion viittaussuhteet näyttävät aiheuttavan laadullisia ongelmia konstruktion tulkinnaissa (mt. 83). Tolvasen (2004) aineisto puolestaan koostuu ylioppilasaineista, joissa hän toteaa infiniittisten konstruktioiden olevan melko harvinaisia niin S2- kuin S1-kirjoitelmissakin – kummassakin osa-aineistossa noin kolmanneksessa aineista on joko yksi tai ei ainuttakaan Tolvasen lauseenvastikkeeksi luokittelemaa infiniittistä konstruktioita (mt. 33).

Käyttöpohjaiseen kielikäsitteeseen kuuluu sisäsyntyisesti ajatus kielenkäytön, kielijärjestelmän, kielitaidon ja kielenoppimisen kontekstikohtaisuudesta ja kielen funktionaalista luonteesta – muodon ja funktion väliset suhteet määrittävät ja ne opitaan osana kielenkäyttöä; monet kielelliset valinnat ovatkin riippuvaisia niiden käyttöympäristöstä (ks. esim. Sinclair 1991, 112). Tekstilaji onkin keskeinen tekijä myös kielen oppimisessa. Swalesin (1990, 58) mukaan tekstilajit ovat kommunikaatiotilanteita, joiden osanottajat jakavat tietyt kommunikatiiviset tavoitteet ja jotka ovat keskenään jossain määrin samankaltaisia rakenteellisesti, tyylillisesti ja sisällöllisesti sekä myös vastaanottajien osalta. Berkenkotterin ja Huckinin (1995, ix) mukaan tekstilajitietoisuus koostuu kielenkäyttäjän kyvystä toimia kielellisesti tyypillisellä 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, kuten 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 niin leksikaalisten kuin kieliopillistenkin piirteiden osalta, että vaihtelu on runsasta yhtäältä erilaisten tekstilajien ja toisaalta 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 myös sen sisäisen variaation osalta (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) esimerkiksi osoittavat, että vielä edistyneetkin kielenoppijat ilmaisevat akateemisissa teksteissä ensikielisiä kielenkäyttäjiä eksplisiittisemmin ja useammin oman näkökulmansa tieteenalasta riippumatta.

S2-tutkimuksen piirissä akateemisen kielenkäytön erityispiirteisiin on perehdytty erityisesti Edistyneiden suomenoppijoiden korpus -tutkimushankkeen piirissä (ks. esim. Siitonen–Niemelä 2011; Ivaska 2014a; 2014b; 2015; Siitonen 2014), joskin eri tekstilajien erityispiirteet ja erot tekstilajien välillä ovat jääneet vähälle huomiolle⁴.

⁴ Ks. kuitenkin vuonna 2015 alkanut tutkimushanke ”Akateeminen suomi kansainvälisen oppijan haasteena”, <http://www.utu.fi/fi/yksikot/hum/yksikot/suomi-sgr/tutkimus/tutkimushankkeet/Sivut/akateeminensuomi.aspx>.

3. Materiaali ja menetelmät

Seuraamme tässä tutkimuksessa ennen kaikkea Siitosen (2014) avaamaa tutkimuslinjaa eli tarkastelemme E-infinitiivikonstruktion todellisia esiintymiä S2- ja S1-kirjoittajien akateemisissa teksteissä ja vertaamme näitä toisiinsa. Laajennamme tarkastelun kattamaan E-infinitiivikonstruktion aktiivi- ja passiivivariantin, minkä avulla voimme havaita käytössä mahdollisesti esiin nousevat konstruktion eri merkitysten käyttöerot eli sen käyttöalan supistumat, laajentumat ja näiden väliset yhteydet. Pyrimme tarkentamaan fokusta yhtäältä sen osalta, missä määrin S2-oppijoiden ja ensikielisten kirjoittajien väliset havaittavat käyttöerot ovat tuottamistilanne- tai tekstilajikohtaisia ja missä määrin erot ovat yhteydessä konstruktion eri merkityksiin. Toisaalta selvitämme, miten ensikielenään suomea puhuvat informantit kokevat tekstilauseiden laadullisen kohdekielenmukaisuuden ja miten yksimielisiä he ovat E-infinitiivikonstruktion laadullisesta kohdekielenmukaisuudesta. 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

Korpusaineistomme on osa Edistyneiden suomenoppijoiden korpusta (LAS2; Ivaska 2014c), ja tarkastelemme kolmea akateemisen kirjoittamisen piiriin kuuluvaa 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 hyvin argumentoituja mielipiteitä opiskelusta sisällöstä. Siinä missä tentit ja esseet ovat luonteeltaan yksityisiä ja tarkoitettu ensisijaisesti opettajan luettavaksi, gradu-käsikirjoitusten lopputuotteena julkaistavat gradut ovat saatavilla myös muiden luettavaksi. Tarkastelemalla näitä tekstilajeja rinnakkain niin S2- kuin S1-aineistossakin on mahdollista tavoittaa havainnoitavan ilmiön luonne ja näin tulkita tarkasteltavien konstruktioiden käyttöön

vaikuttavia tekijöitä ja samanaikaisesti myös infinitiivikonstruktioiden ja tekstilajin suhdetta kielen kompleksisuuteen. Taulukko 1 kuvaa käyttämämme korpusaineiston koon.

Taulukko 1. Korpusaineiston koko ja jakauma osa-aineistoittain.

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

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 kielellisin metatiedoin lemموjen, sanaluokkien, morfologisten muotojen ja syntaktisten funktioiden osalta. Koodaus on kohdistettu sanatasolle, joten yhdellä sanalla saattaa olla samanaikaisesti useita morfologisia koodeja. (Ks. tarkemmin Ivaska 2014c.)

Korpusanalyysimme on tutkimusotteeltaan 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, minkä lisäksi vertailemme eri osa-aineistoja keskenään. Tältä osin tutkimuksemme seuraa oppijankielen korpustutkimuksessa yleistä kontrastiivista välikielianalyysia (engl. *Contrastive Interlanguage Analysis* – CIA; Granger 1996; 2015). Vertailu on monitasoista, sillä vertaamme toisiinsa sekä S2- ja S1-aineistoja että eri tekstilajeja.

Tarkastelu on määrällistä, sillä lähtökohtanamme ovat morfologisesti määriteltyjen E-infinitiivikonstruktioiden käyttöfrekvenssit. Käyttöfrekvenssien ja niihin perustuvien tulkintojen lisäksi perehdymme kuitenkin myös konstruktion laadullis-funktionaaliseen vaihteluun. Pyrimme näin vastaamaan CIA-tutkimusta kohtaan esitettyyn kritiikkiin siitä, että menetelmää soveltavat tutkimukset keskittyvät liikaa yksittäisten kielenpiirteiden frekvenssieroihin ja jättävät huomiotta niiden käyttöympäristön ja sen sisällään pitämät

kielelliset ja kielenulkoiset muuttujat (ks. esim. Gries–Deshors 2014). Ennen kaikkea tarkastelemme sitä, miten konstruktioita käytetään eri tekstilajeissa, mitkä ovat sille tyypillisiä leksikaalisia representaatioita ja mitä funktioita se tyypillisesti ilmaisee. Käytännössä korpusanalyysimme käsittää seuraavat vaiheet:

1. Kootaan LAS2:sta kooltaan tasapainoinen otos tarkasteltavien tekstilajien tekstejä siten, että tekstejä on S2- ja S1-aineistossa kussakin tekstilajissa yhtä paljon;
2. Haetaan aineistosta kaikki inessiivisijaisen E-infinitiivikonstruktion esiintymät;
3. Tehdään tilastollinen vertailu konstruktion käyttöfrekvensseistä S2- ja S1-aineistojen sekä eri osa-aineistojen välillä;
4. Käydään läpi kaikki konstruktioesiintymät kontekstissaan sekä tulkitaan ja koodataan aineistoon niiden ensisijainen merkitys. Isoa suomen kielioppia (VISK § 543–§ 545) seuraten muuttujan mahdolliset arvot ovat aikasuhde, syysuhde ja ehtosuhde;
5. Tarkastellaan konstruktion eri merkitysten käyttöaloja, skemaattisuutta ja leksikaalista jakaumaa eri osa-aineistoissa.

Konstruktioiden merkitysten luokittelu on väistämättä aina tulkinnanvaraista, ja nyt käsillä olevan kaltaisen monimerkityksisen tai merkitykseltään haalistuneen konstruktion tapauksessa tämä on vielä ilmeisempää. Esimerkiksi kausaalisuus ja konditionaalisuus ovat jo itsessään jatkumomaisia ilmiöitä. Noordman ja de Blijzer (2000) osoittavatkin kausaalisten ilmausten merkityksen ymmärtämisen riippuvan samanaikaisesti niin sisällön ja episteemisten kytkösten välisestä suhteesta, syyn ja vaikutuksen (engl. *cause and effect*) välisestä lineaarisesta järjestyksestä lauseessa kuin syyn ja vaikutuksen käsitteellisestä järjestyksestäkin. Nyt käsillä oleva konstruktio kuvaa prototyyppisesti aikasuhdetta – myös kaikissa aineistomme konstruktioissa temporaalinen tulkinta on mahdollinen, ja esimerkiksi kiteytyneiden *mennessä* ja *kuluessa* -tapauksen kohdalla se lienee ainoa tulkinta. Muissa tapauksissa olemme valinneet ensisijaiseksi tulkinnaksi kontekstin parhaiten tukeman vaihtoehdon. Käytimme tulkinnan apuna kutakin merkitystä vastaavaa parafrasaa seuraavasti: konditionaalisen parafrasina oli käytössä 'sillä ehdolla, että', kausaalisen parafrasina oli 'siitä syystä, että' ja ne tapaukset, joihin kumpikaan parafrasi ei sopinut, tulkitsimme ensisijaisesti temporaalisiksi. Epäselvimmät tapaukset olisi nähdäksemme voinut jättää käsittelemättä tai soveltaa luokittelussa epäselvät-kategoriaa tai pyrkiä löytämään todennäköisimmät ratkaisut. Olemme päätyneet viimeiseen vaihtoehtoon keskustelemalla

monitulkintaisimmista tapauksista yhdessä ja valinneet tulkin, joka tyydyttää kumpaakin kirjoittajaa.

Kaikki hakutulokset on tallennettu LAS2-hakutyökalujen avulla csv-muotoon. Tilastolliset analyysit on tehty R-ohjelmointiympäristössä (R Core Team 2015). Käyttöfrekvenssien tilastollisessa tarkastelussa olemme käyttäneet tekstikohtaisia suhteellisia frekvenssejä eli esiintymien määrää kussakin tekstissä tuhatta sanetta kohti. Esiintymisfrekvensseissä havaittujen tilastollisten erojen testinä olemme käyttäneet Mann-Whitneyn U-testiä, joka on luonteeltaan epäparametrinen eikä näin ollen herkkä esiintymien epäsymmetrisille jakaumille. Konstruktion skemaattisuutta tarkastelemme vertaamalla toisiinsa 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 siinä esiintyvän piirteen – kuten tietyn lekseemin – välisen yhteyden vahvuutta vertaamalla näitä kummankin kokonaismääriin aineistossa. Menetelmä on kehitetty alun perin funktioltaan samankaltaisten konstruktioiden leksikaalis-semanttisten preferenssien tarkasteluun (Gries–Stefanowitch 2004), minkä lisäksi sitä on käytetty tämän nyt käsillä olevan tutkimuksen tapaan saman konstruktion leksikaalis-semanttisten käyttöerojen tarkasteluun eri osa-aineistoissa (ks. esim. Gries–Hilpert 2008; Ivaska 2015).

3.2. Kyselytutkimukset

Kyselytutkimuksemme seuraavat aiempien käyttöpohjaisessa viitekehyksessä tehtyjen 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 kyselyssä olleista virkkeistä on täysin kohdekielenmukaisia, kun taas osassa lauseista on jotakin yllättävää nimenomaan tutkitun infinitiivikonstruktion kannalta.

Kyselytutkimus 1

Kyselytutkimuksessa 1 vastaajia pyydettiin arvioimaan yhteensä 44 suomenkielisen LAS2-korpuksesta kerätyn virkkeen oikeakielisyyttä viisiportaisella asteikolla niin, että (1) on vastaajan mielestä yksiselitteisesti epäkieliopillinen ja (5) vastaavasti yksiselitteisesti kieliopillinen. Virkkeistä kymmenessä esiintyi tutkimuskohteena oleva E-infinitiivikonstruktio ja muissa lauseissa jokin muu infinitiivikonstruktio. Tällä tavoin pyrimme välttämään ns. priming-ilmiötä eli vastaajien virittymistä yhden tietyn konstruktion tulkintaan. Vastaajille ei kerrottu etukäteen, mihin esitettyjen virkkeiden ominaisuuteen heidän tuli kiinnittää huomiota. Arvioiden jälkeen vastaajia pyydettiin järjestämään samat kymmenen tutkittavaa virkettä alenevaan järjestykseen kieliopillisimmasta epäkieliopillisimpaan. Tarkastellut konstruktiot 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** ääniteitä 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 lauseiden saamien oikeakielisyyсарvioiden keskiarvoa ja sitä, miten yhdenmukaisia eri vastaajien arviot ovat keskenään ja missä määrin vastaajien esittämät virkkeiden kieliopillisuusjärjestykset poikkeavat toisistaan. Lisäksi tarkastelemme sitä, miten nämä vastaukset suhteutuvat konstruktion semanttisiin funktioihin ja niiden käyttöaloihin eri osa-aineistoissa. Kyselyaineisto on kerätty Turun alueella kevään 2015 aikana 23:lta suomea ensikielenään puhuvalta informantilta. Vastaajat olivat akateemisesti koulutettuja ja heistä 19 joko opiskeli tai työskenteli kielen parissa.

Kyselytutkimus 2

Kyselytutkimus 2 täydentää kyselytutkimusta 1. Vastajia pyydettiin järjestämään samat kymmenen tutkittavaa virkettä (esimerkit 6–15) alenevaan järjestykseen kieliopillisimmasta epäkieliopillisimpaan. Tämän jälkeen vastaajille esitettiin samat virkkeet sekä aktiivisessa että passiivisessa muodossa ja heitä pyydettiin arvioimaan virkkeiden oikeakielisyyttä viisiportaisella asteikolla niin että (1) on vastaajan mielestä yksiselitteisesti epäkieliopillinen ja (5) vastaavasti 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 virkkeet viisiportaisella asteikolla niin, että (1) on vastaajan mielestä yksiselitteisesti mahdoton tai järjetön ja (5) puolestaan kaikin puolin mahdollinen ja järkevä. Lopuksi vastaajia pyydettiin kommentoimaan alkuperäisiä virkkeitä (esimerkit 6–15) sen kannalta, mikä niissä on heidän mukaansa kohdekielenvastaista. Kyselyaineisto on kerätty Turun alueella syksyn 2016 aikana, ja siihen vastasi 35 informanttia, joista 25 puhuu ensikielensä suomea ja 10 jotakin muuta kieltä. Kaikki vastaajat olivat akateemisesti koulutettuja ja he opiskelivat kieliaineita. Ensikielensä jotakin muuta kuin suomea puhuvia vastaajia voidaan pitää korpusaineistomme 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 / 1000 sanetta) on aktiivisia ja 113 (0,52 / 1000 sanetta) on passiivimuotoisia. Konstruktiosta 129 on S2-aineistossa ja 226 on S1-aineistossa. Taulukko 2 havainnollistaa esiintymämääriä aineistoittain 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 ylipäätään verraten harvinainen – 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. Lisäksi on merkillepantavaa, että erityisesti S2-aineistossa passiivikonstruktiota esiintyy lähes yksinomaan graduissa.

Aktiivi- ja passiivivariantti ovat kumpikin S2-aineistossa harvinaisempia kuin S1-aineistossa, ja ero on molemmissa tapauksissa tilastollisesti merkitsevä⁵. Osa-aineistojen tarkastelu osoittaa lisäksi, että S2- ja S1-aineistojen välinen frekvenssiero on tilastollisesti merkitsevä tenteissä ja esseissä passiivikonstruktioiden osalta ja ainoastaan tenteissä aktiivikonstruktioiden osalta. Graduissa konstruktioita käytetään itse asiassa S2-aineistossa S1-aineistoa enemmän, joskaan ero ei ole tilastollisesti merkitsevä. Toisin sanoen konstruktioiden käyttö ei välttämättä erota yleisesti ensikielenään suomea puhuvia kirjoittajia ja suomen taidoiltaan edistyneitä ei-ensikielisiä kirjoittajia toisistaan, vaan ero perustuu ainakin tämän tutkimuksen aineiston valossa myös käsillä olevaan tekstilajiin – huolimatta siitä, että nyt käsillä olevia tekstilajeja voitaneen pitää monin tavoin toistensa kaltaisina.

⁵ Tilastollisen merkitsevyyden tulkinta on tässä artikkelissa tehty tasolla $p < .05$.

4.2. Funktionaalinen ja leksikaalinen vaihtelu

Aineiston 355 konstruktioesiintymästä noin 46 % on funktioltaan ensisijaisesti ajallista suhdetta, noin 31 % on ehtosuhdetta ja noin 21 % kausaalista vaikutussuhdetta kuvaavia. Loput esiintymät kuvaavat rinnasteisuutta, tai niiden merkitys on tulkittavissa epäselväksi. Koska rinnastusta ilmaisevia tai merkitykseltään epäselviä tapauksia on niin vähän, ne on jätetty määrällisen tarkastelun ulkopuolelle. Esimerkki 16 kuvaa esiintymää, jonka olemme tulkinneet ensisijaisesti ajallista suhdetta 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 ensikielät ovat englanti, komi, liettua, ruotsi, tšekki, venäjä ja viro. las2-vtt02vert002)*
- (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-vtt01vert127)
- (18) *Luonnehtimisen kannalta on myös ollennaista onko kieli ylhäällä tai alhaalla. Diagrammassa näkee, että i, y ja ö ovat suppeita vokaaleja, eli kielenkärki niitä ääntäessa on ylhaalla. e, ö ja o ovat puolisuppeita ja ä, a laveita.* (las2-14tt01te01lo00)

Funktiot jakautuvat aktiivi- ja passiivikonstruktioissa hyvin eri tavoin, sillä aktiivisista konstruktioista noin 58 % (n = 142) ilmaisee aikasuhdetta, 12 % (n = 28) ehtosuhdetta ja noin 28 % (n = 71) kausaalisuhdetta, kun taas passiivikonstruktioista noin 19 % (n = 22) ilmaisee aikasuhdetta, noin 73 % (n = 83) ehtosuhdetta ja noin 7 % (n = 8) kausaalisuhdetta. Taulukko 3 kuvaa näiden funktioiden esiintymistä osa-aineistoittain.

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	31	19	20	71	71
	passiivi	11	2	5	2	2	0	18	4
KAUSAALI	aktiivi	29	7	8	8	11	8	48	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

Aktiivikonstruktio on passiivivarianttia selvästi yleisempi kahden teon ajallista suhdetta tai tekojen keskinäistä vaikutussuhdetta ilmaisemassa, kun taas passiivikonstruktio on selvästi aktiivista varianttia yleisempi tekojen välistä ehtosuhdetta ilmaisemassa. Aktiivin edellä todettu 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 aiemmin todettu tilastollisesti merkitsevä käyttöero tenttivastauksissa heijastuu niin ikään kaikissa konstruktion funktioissa, joskin se on selvin konstruktion yleisimmässä funktiossa – ehtosuhteen ilmaisijana. Passiivisten konstruktioiden osalta S2- ja S1-aineistojen välinen käyttöero näyttää siis yhtäältä olevan tekstilajikohtainen mutta toisaalta myös merkityskohtainen – passiivivarianttia käytetään ilmaisemaan nimenomaan ehtosuhdetta, mikä puolestaan näyttää aineiston valossa tyypilliseltä nimenomaan 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 erikseen että koko aineistossakin. Toisin sanoen 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. Tämän voi tulkita tarkoittavan sitä, että konstruktioiden prototyyppinen käyttö on S2-aineistossa yleisempää ja tekstilajikohtainen funktionaalinen vaihtelu vähäisempää. Tenteissä havaittava passiivisten ehtoa ilmaisevien konstruktioiden poikkeava jakauma ei täysin sovi tähän tulkintaan, mutta poikkeus saattaa johtua siitä, että ehtoa ilmaiseva funktio on ylipäättään vasta kehkeytymässä osaksi S2-kirjoittajien spontaanissa tuottamistilanteessa käyttämää kielirepertoaria.

Tämän aineiston valossa on mahdotonta sanoa, minkälainen näiden kahden muuttujan välinen vaikutussuhde on luonteeltaan tai suunnaltaan. Onko kyseessä siis edistyneiden kielenoppijoiden 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 huomioonottaminen 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 leksikaalisen vaihtelun tarkastelu 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- kuin ehtosuhdettakin ilmaisemaan käytettyjä kommunikaatioverbejä, kuten *puhua*, *arvioida* ja *lukea*, ja geneerisiä, 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 funktionaalisesikin vaihtelusta.

Aikasuhdetta kuvaavat verbit ovat suhteessa yleisempiä S2-aineistossa, kun taas S1-aineistossa korostuvat enemmän 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 antaisi olettaa (n. 8 esiintymää) 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-infinitiivikonstruktioiden vaihtelua on vähemmän ja että konstruktion prototyyppiset merkitykset kattavat suuremman osan kaikista esiintymistä.

4.3. Oikeakielisyyssarvionnit kyselytutkimuksissa

E-infinitiivikonstruktioiden oikeakielisyyssarvioissa nousee esiin muutamia varteenotettavia seikkoja. Taulukossa 6 esimerkit on järjestetty hyväksyttävimmästä vähiten hyväksyttävään niiden oikeakielisyyden järjestytehtävissä saamien arvioiden keskiarvojen perusteella. Taulukko havainnollistaa siis oikeakielisyyssarvioiden keskiarvoja sekä järjestytehtävissä saatujen vastausten keskiarvoja.

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 kielioppia 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>Komponenttianalyysia 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 ilmapirta 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 ääniteitä 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 tulkkaukspalvelu kaikkien jäsenvaltioiden virallisten kielten ollessaan myös työkieliä.</i>	2.7	6.65=10	7.2=10	6=9

Tarkastellut kymmenen esimerkkiä jakautuvat arvioiden kannalta kolmeen ryhmään, jotka toistuvat sekä esimerkikohtaisissa oikeakielisyyсарvioissa että S1-arvioijien tekemissä listoissa esimerkkien suhteellisesta oikeakielisyydestä. Yksi esimerkki (taulukon 6 rivi 1) on luokiteltu selvästi muita oikeakielisemmäksi, ja se erottuu myös oikeakielisyyjärjestykseen perustuvassa listauksessa selvästi muista esimerkeistä. Kuusi muuta esimerkkiä (taulukon 6 rivit 2–7) muodostavat toisen ryhmän, joiden arviot ovat keskenään samankaltaisia ja joiden keskinäinen järjestykseen perustuva luokitus 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.⁶ Mielenkiintoista on, että nämä ryhmät hahmottuvat taulukon riviä 7 lukuun ottamatta samankaltaisina sekä kyselytutkimuksessa 1 että kyselytutkimuksen 2 S1-aineistossa, kun taas kyselytutkimuksessa 2 kerätyssä S2-aineistossa arvioitujen lauseiden keskinäinen järjestys on hyvin erilainen eikä se noudata edellä kuvattua ryhmittymistä.

⁶ Kuten nämä arviot osoittavat, ensikielisetkään arvioijat eivät koskaan ole täysin yksimielisiä. Nyt käsillä olevan konstruktion käytössä ja sen hallinnassa on vaihtelua ja horjuntaa myös ensikielisten kielenkäyttäjien piirissä (Kurki 2016).

Tulkitsemme tulosta siten, että ensikieliset arvioijat ovat verraten yksimielisiä siitä, että rivin 1 esimerkki on kohdekielenmukainen, rivien 2–7 esimerkeissä on jotakin yllättävää mutta ne koetaan kuitenkin verraten 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 sanajärjestyksensä ja *mieluimmin*-sanan vuoksi, ei infinitiivin takia.⁷ Samat ryhmittymät eivät kuitenkaan nouse esiin S2-aineistossa, eli suomea toisena kielenään puhuvat vastaajat eivät jaa samaa käsitystä lauseiden kohdekielenmukaisuudesta.

Ensikieliset arvioijat kokevat passiivikonstruktion aktiivikonstruktiota kohdekielenmukaisemmaksi kaikissa tarkastelluissa virkkeissä taulukon 6 rivin 10 esimerkkiä lukuun ottamatta – passiivimuotojen arvioiden keskiarvo on kussakin tapauksessa viisiportaisella asteikolla yli 3.8. Aktiivimuodoista kohdekielenmukaisimmaksi on arvioitu rivin 6 *Ääntäessä ilmavirta pääsee kurkunpäästä esteettä ulos*, ja sen saamien arvioiden keskiarvo on 3.01. Rivin 10 esimerkivirke on todettu selvästi kohdekielenvastaiseksi niin aktiivi- kuin 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 kohdekielenmukaiseksi kolmessa tapauksessa (taulukon rivit 3, 4, 8) aktiivinen muoto on koettu selvästi passiivista muotoa kohdekielenmukaisemmaksi ja kahdessa tapauksessa (rivit 1 ja 7) passiivinen muoto on koettu aktiivista muotoa kohdekielenmukaisemmaksi.

Kyselytutkimuksen 2 suomea ensikielenä puhuvat vastaajat ovat lisäksi verraten yksimielisiä siitä, mikä kussakin lauseessa on epäodotuksenmukaista tai kohdekielenvastaista. Kuten taulukko 7 kuvaa, tarkastelluista kymmenestä virkkeestä kuudessa (rivit 2, 3, 5, 6, 8 ja 9) ensikieliset arvioijat kokevat ongelmaksi E-infinitiivin pääluokan tai ainakin E-infinitiivin ylipäättään. Kahdessa tapauksessa (rivit 4 ja 10) ongelman katsotaan liittyvän omistusliitteeseen ja kahdessa tapauksessa (rivit 1 ja 7) ei valtaosan mielestä ole mitään

⁷ 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 ilmipassiivin taustalla olevien tekijöiden tulkinta aiheuttaa sumeutta. Näitä kysymyksiä ei voi kuitenkaan pohtia enempää tämän artikkelin rajoissa.

kohdekielenvastaista. On lisäksi huomionarvoista, että useassa tapauksessa iso osa ensikielisistä arvioijista ei ehdota minkäänlaisia korjausta huolimatta siitä, että arvioinneissa lauseissa arvioitiin olevan jotakin kohdekielenvastaista. Suomea toisena kielenä puhuvat arvioijat esittävät vielä vähemmän korjausehdotuksia; valtaosa heistä ei tarjoa minkäänlaisia korjauksia tarkasteltuihin virkkeisiin. Kun nämä arvioijat ehdottavat jotakin korjausta, ne kohdistuvat useimmiten infinitiiviin tavalla tai toisella: he esimerkiksi ehdottavat sijalle jotakin toista infinitiivikonstruktiota, kuten MA-infinitiiviä. Omistusliitteen käyttöön liittyviä korjausehdotuksia ei ole kuin yksi kappale.

Taulukko 7. Kymmenen E-infinitiivikonstruktion korjausehdotukset.

	ESIMERKKI	INFINITIIVI+ PASSIIVI		INFINITIIVI		INFINITIIVI + 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 ilmvirta 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 tulkkauksen palvelu kaikkien jäsenvaltioiden virallisten kielten ollessaan myös työkieliä.</i>	1	0	3	0	12	0	7	3	2	7

Rivin 7 virkkeeseen *Maija vastaa kysyttäessä, että hän asuisi kyllä mieluummin Ahvenanmaalla* liittyviä korjausehdotuksia on syytä tarkastella erikseen sekä ensikielisten että toisella kielellä toimivien arvioijien osalta, sillä kumpikin ryhmä piti kyselytutkimuksessa 2 virkettä selvästi kohdekielenvastaisena. Korjausehdotuksia lauseelle ehdotetaan kuitenkin hyvin vähän. Tarkempi tarkastelu paljastaakin, että virkettä koskevat korjaukset liittyvät useimmiten joko ensimmäisen lauseen sanajärjestykseen tai *mieluummin*-sanankäyttöön – eivät infinitiivin käyttöön.

Sekä esimerkkikohtaiset että järjestykseen perustuvat arviot muodostavat samankaltaisen ryhmittymisen, mikä vahvistaa ajatusta siitä, että ensikieliset kielenkäyttäjät kokevat esimerkkilauseiden kieliopillisuuden keskenään samalla tavalla. Tarkastelemme seuraavaksi kolmea vähiten kohdekielisenä pidettyä esimerkkiä yksityiskohtaisemmin.

- (19) *Kantakielestä nykysuomeen **siirtyessä** monta konsonanttia on kadonnut.*
(las2-20tt01te01)
- (20) ***Opettaessa** äänneitä oppijoiden täytyy muistaa, että suomi on vokaalirikas kieli.* (las2-21tt01te07)

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 pääluokkaan – niin, että 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 siis ilmaisisi sitä ehtoa, että tarkastelun fokus siirretään kantakielestä nykysuomeen. Esimerkissä 20 muistamisen ehtona olisi opetustilanne. Esimerkin 19 kohdalla kaikki arvioijat pitivät ehtosuhde- ja aikasuhde-parafraseja suunnilleen yhtä järkevinä (S1: ehtosuhde 3.3, kausaalisuhte 1.2 ja aikasuhde 3.6; S2: ehtosuhde 3.1, kausaalisuhte 1.7 ja aikasuhde 3.6⁸) ja esimerkin 20 kohdalla aikasuhde-parafrasi arvioitiin järkevimmäksi, joskin myös ehtosuhde koettiin täysin mahdolliseksi (S1: ehtosuhde: 3.6, kausaalisuhte 1.3 ja aikasuhde 4.6; S2: ehtosuhde 3.1, kausaalisuhte 2.0 ja

⁸ Myös näissä arvioissa käytössä oli viisiportainen asteikko, jossa (5) vastaa täysin järkeenkäypää ilmausta ja (1) täysin järjetöntä ilmausta.

aikasuhde 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, mutta tämä ei vaikuttane konstruktion merkitykseen. Tulkinnan epäselvyys saattaakin olla yksi keskeinen tekijä oikeakielisyyсарvioinnin 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.* (las2-35tt01te02)

Esimerkin 21 konstruktion merkitys on kaukana aikasuhtetulkinnasta, ja se lähenee luonteeltaan kausaalisuhteen ilmaisemista niin, että esimerkin parafraasi 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, aikasuuhde 2.9; S2: ehtosuhde 2.7, kausaalisuhde 4.1, aikasuuhde 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. Näin ollen tapaus voidaan nähdä esimerkkinä samanaikaisesta useiden kielellisesti epätyypillisten piirteiden lankeamisesta yhteen, minkä on esitetty lisäävän kohdekielenvastaisten ilmausten mahdollisuutta (ks. Ivaska 2011). Esimerkin 21 kohdalla voidaan ajatella, että kompleksisen ja verraten harvinaisen konstruktion (-essA-konstruktio) epätyypillisen merkityksen (kausalisuhde) kanssa lankeaa yhteen kielenpiirre (possessiivisuffiksi), jonka tyypillinen käyttö rajautuu tiettyihin yhteyksiin (kirjoitettu kieli) ja jota käytetään tyypillisesti jonkin toisen sanaluokan (substantiivi) yhteydessä.

5. Päätelmät, tulkinnat ja yhteenveto

Analyysin perusteella esitämme tutkimuskysymyksiimme seuraavia vastauksia: inessiivisijaisen E-infinitiivikonstruktion käyttö on kaikkiaan melko harvinaista – konstruktiota esiintyy alle puolessa tutkituista teksteistä. Konstruktion aktiivimuotoinen variantti on yleisempi niin S2- kuin S1-aineistossakin, joskin S2-aineistossa ero on S1-aineistoa suurempi. Tutkituista tekstilajeista konstruktion käyttö erottaa S2- ja S1-aineistoja toisistaan tilastollisesti merkitsevästi tenteissä niin aktiivi- kuin passiivikonstruktioidenkin osalta ja esseissä passiivikonstruktioiden osalta. Kaikissa näissä tapauksissa konstruktiio on yleisempi S1-aineistossa. Vajaa puolet konstruktiioista ilmaisee ensisijaisesti ajallista suhdetta, noin kolmannes ehtosuhdetta ja noin viidennes kausaalista vaikutussuhdetta. Nämä funktiot jakautuvat aktiivi- ja passiivikonstruktiioihin epätasaisesti siten, että aikasuhte on selvästi yleisin merkitys aktiivisille konstruktiioille, kun taas passiivisista konstruktiioista valtaosa ilmaisee ehtosuhdetta. Tekstilajien välisistä eroista vahvimpana nousee esiin se, että ehtosuhteen ilmaiseminen näyttää ennen kaikkea tenteille tyypilliseltä. S2- ja S1-aineistot eroavat toisistaan ennen kaikkea siten, että S2-aineistossa konstruktion tyypillisin merkitys – aktiivisissa tapauksissa aikasuhte ja passiivisissa tapauksissa ehtosuhte – kattaa suuremman osan konstruktion kaikista esiintymistä. Leksikaalisen variaation tarkastelu tukee tätä tulkintaa, sillä ensisijaisesti aikasuhdetta ilmaisevat verbit ovat suhteellisesti yleisempiä S2-aineistossa, kun taas yleisemmät, myös ehtosuhdetta ilmaisevat verbit ovat suhteellisesti yleisempiä S1-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, mutta uskomme, että havaintomme merkityssuhteen ja pääluokan välisistä suhteista samoin kuin tyypillisimpien merkitysten yleisyydestä ovat joka tapauksessa valideja.

Ensikielisten suomenoppijoiden tekemät inessiivisijaisten E-infinitiivikonstruktioiden oikeakielisyydsarviot osoittavat, että arvioijat tunnistavat edistyneiden suomenoppijoiden tuottamien konstruktioiden poikkeamisen kohdekielestä ja kokevat konstruktioiden oikeakielisyyden melko samalla tavalla. Kymmenestä tarkastellusta esimerkistä nousee esiin kohdekielisyyden kannalta kolme erilaista ryhmää: yksi esimerkki koettiin lähes yksiselitteisesti kohdenkielenmukaiseksi, kuusi esimerkkiä vaikutti osin kohdekielenvastaisilta ja kolme esimerkkiä koettiin selvästi kohdekielenvastaisiksi. Suomea toisena kielenä

oppineiden arvioijien vastauksista nämä ryhmät eivät nouse millään lailla esiin. Ensikieliset arvioijat olivat lisäksi hyvin yksimielisiä siitä, että ongelmat liittyivät useimmiten 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 niiden merkitys jää monitulkintaiseksi. Kyselytutkimusten oikeakielisyyssarvioita tarkasteltaessa on syytä muistaa, että emme ole tarkastelleet tulosten tilastollista merkitsevyyttä – osin siksi, että tällainen tarkastelu ei nähdäksemme merkittävästi vahvistaisi jo nyt melko selviä havaittuja tendenssejä, ja osin myös siksi, että etenkin suomea toisena kielenä oppineiden arvioijien vastauksia on 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 konstruktioit viittanevat kumpikin 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 S2-aineistossa vastaa monilta osin aiemmassa tutkimuksessa (ks. esim. Lauranto 1997; Martin ym. 2010; Mustonen 2015) havaittua konstruktioiden konkreettisten merkitysten vähittäistä vaihtelun kasvua ja abstraktistumista. Erot funktionaalisessa distribuutiossa kuvannevat tällaista osittaista tai meneillään olevaa oppimista, sillä S2-aineistossa korostuvat S1-aineistoa enemmän konstruktion eri varianttien prototyypisemmät muodot ja oikeakielisyyssarvioissa 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 dynaaminen ja monitahoinen perusluonne (ks. esim. Larsen-Freeman 1997; Larsen-Freeman–Cameron 2008) ja siihen vaikuttavat erilaiset muuttujat. Tässä tutkimuksessa tarkastelimme konstruktioita sen muotovariaation, leksikaalisen variaation, funktionaalisen variaation ja tekstilajien variaation osalta, joista kukin valottaa nähdäksemme konstruktion käyttöä ja sen

hallintaa osin eri tavalla ja joista yhdenkin poisjättäminen yksinkertaistaisi ja osin myös vääristäisi saatavaa kokonaisuutta. Tämä on erityisen relevanttia kompleksisissa konstruktioiden, joihin vaikuttaa niin muoto- kuin merkityspiirteidenkin kannalta useita samanaikaisia mekanismeja. Toinen tutkimuksen metodologinen vahvuus on korpusmenetelmien yhdistäminen oikeakielisyyssarvioihin. Tieteellisessä keskustelussa on kaivattu korpus- 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 puitteissa varsinaisesti paneudu kohdekielen mukaisuuden, 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 välttää samanaikaisesti vertailuharha eli ajatus monoliittimaisesta staattisesta ja vaihteluttomasta kohdekielen järjestelmästä, johon oppijankieltä verrataan.

Konstruktion eri muotojen ja merkitysten käyttö on myös tekstilajikohtaista, sillä jopa tämän tutkimuksen aineistossa olleiden, keskenään verraten samankaltaisten tekstilajien välillä näyttää olevan selviä eroja konstruktion varianttien ja niiden funktioiden distributiivissa. Käyttöpohjaisesta näkökulmasta on perusteltua kysyä, missä määrin konstruktion eri merkitysten oppiminen edellyttää niille tyypillisten käyttökontekstien tarvetta ja altistumista näille konteksteille sekä näille käyttöympäristöille tyypillisten tekstilajien tuntemusta ja niiden 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 ehto ilmaiseva funktio on vasta kehkeytyvässä osaksi S2-kirjoittajien spontaanissa tuottamistilanteessa – tässä tapauksessa tentteissä – käyttämää kielirepertoaria, vaikka sen käyttö on muissa tarkastelluissa tekstilajeissa lähellä S1-kirjoittajien käyttöä. Tentit eroavat graduista ja esseistä niin suunnittelun ja huolittelemisen kuin yksityisyyden ja odotettujen lukijoidenkin osalta. S2- ja S1-aineistojen välinen ero voidaan siis nähdä konstruktion käytön osittaisen oppimisen lisäksi myös tekstilajin ja sille tyypillisen rekisterin osittaisena oppimisena. Nähdäksemme konstruktiokohtaista oppimista

tutkittaessa olisikin syytä ottaa aina huomioon ilmaistujen merkitysten monipuolistumisen (vrt. Seilonen 2013; Kajander 2013) lisäksi myös käyttökotekstien monipuolistuminen ja kullekin käyttökotekstille tyypillisen kielenkäytön tavat. Nämä näkökulmat kuvaavat luonnollisesti saman ilmiön eri puolia, mutta on syytä muistaa, että käyttöpohjaisten mallien hengessä kaikki kielenkäyttö ja kielenkäytön tyypillisuus on kontekstikohtaista ja että näin ollen konstruktoiden oppiminen on siis aina myös tekstilajien oppimista.

Lähteet

- AIJMER, K. (toim.) 2009: *Corpora and Language Teaching*. Studies in Corpus Linguistics 33. Amsterdam: John Benjamins.
- BARDOVI-HARLIG, K. – Z. DÖRNYEI 1998: Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. – *TESOL Quarterly* 32(2), 233–259.
- BERKENKOTTER, C. – T. N. HUCKIN 1995: *Genre knowledge in disciplinary communication*. Hillsdale (NJ): Lawrence Erlbaum.
- BIBER, D. 2006: *University Language. A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins. DOI: <http://dx.doi.org/10.1075/scl.23>.
- BIBER, D. – S. CONRAD 2009: *Register, Genre, and Style*. Cambridge: Cambridge University Press. DOI: <http://dx.doi.org/10.1017/CBO9780511814358>.
- BIBER, D. – S. CONRAD – R. REPPEN – P. BYRD – M. HELT 2002: Speaking and writing in the university: A multidimensional comparison. – *TESOL Quarterly* 36(1), 9–48.
- BIBER, D. – S. CONRAD – R. REPPEN – P. BYRD – M. HELT – V. CLARK – V. CORTES – E. CSOMAY – A. URZUA 2004: *Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus*. ETS TOEFL Monograph Series 25. Princeton (NJ): Educational Testing Service.
- BULTÉ, B. – A. HOUSEN 2012: Defining and operationalizing L2 complexity – A. Housen, F. Kuiken ja I. Vedder (toim.), *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA*, 21–46. Amsterdam: John Benjamins.
- BYBEE, J. – D. EDDINGTON 2006: A usage-based approach to Spanish verbs of 'becoming'. – *Language* 82(2), 323–355.
- BYBEE, J. – S. 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.
- CALLIES, M. 2013: Agentivity as a determinant of lexico-syntactic variation in L2 academic writing. – *International Journal of Corpus Linguistics* 18(3), 357–390. DOI: 10.1075/ijcl.18.3.05cal.
- CALLIES, M. – E. ZAYTSEVA 2013: The Corpus of Academic Learner English (CALE) – A new resource for the study and assessment of advanced language proficiency. – S. Granger, G. Gilquin ja F. Meunier (toim.), *Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Corpora and Language in Use – Proceedings* 1, 49–59. Louvain-la-Neuve: Presses universitaires de Louvain.
- CADIerno, T. – S. ESKILDSEN (toim.) 2015: *Usage-Based Perspectives on Second Language Learning*. Berlin: Mouton de Gruyter.
- CHEN, M. 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, S. 2005: Corpus linguistics and L2 teaching. – E. Hinkel (toim.), *Handbook of research in second language teaching and learning*, 393–409. Mahwah, NJ: Lawrence Erlbaum.
- CONRAD, S. 2011: Variation in corpora and its pedagogical implications. – V. Viana, S. Zyngier ja G. Barnbrook (toim.), *Perspectives on corpus linguistics* 47–62. Amsterdam: John Benjamins.
- COTTON, F. 2010: Critical thinking and evaluative language use in academic writing: A comparative cross-cultural study. – G. Blue (toim.). *Developing Academic Literacy*. Oxford: Peter Lang, 73–85.
- DĄBROWSKA, E. 2004: *Language, Mind and Brain. Some Psychological and Neurological Constraints on Theories of Grammar*. Edinburgh: Edinburgh University Press.
- DĄBROWSKA, E. 2008: Questions with long-distance dependencies: A usage-based perspective. – *Cognitive Linguistics* 19(3), 391–425.
- EKLUND, M. 1996. *Temporaaliset sivulauseet ja lauseenvastikkeet ulkomaalaisopiskelijoiden kirjoitelmissa*. Sivulaudaturtutkielma. Helsingin yliopiston suomen kielen laitos.
- ESKOLA, S. 2002. *Syntetisoivat rakenteet käännessuomessa. Suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmällä*. Joensuun yliopiston humanistisia julkaisuja 30.
- GOLDBERG, A. E. 1995: *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- GOLDBERG, A. E. 2006: *Constructions at Work – The Nature of Generalization in Language*. Oxford: Oxford University Press.
- GOLDBERG, A. E. – D. CASENHEISER 2008: Construction learning and second language acquisition. – P. Robinson ja N. C. Ellis (toim.), *Cognitive Linguistics and Second Language Acquisition*, 197–215. Amsterdam: John Benjamins.
- GOLDBERG, A. E. – D. CASENHEISER – N. SETHURAMAN 2004: Learning argument structure generalizations. – *Cognitive Linguistics* 15(3) 289–316.
- GRANGER, S. 1996: From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. – K. Aijmer, B. Altenberg ja M. Johansson (toim.), *Languages in Contrast*, 37–51. Lund: Lund University Press.
- GRANGER, S. 2015: Contrastive Interlanguage Analysis: A Reappraisal. – *International Journal of Learner Corpus Research* 1(1), 7–24.
- GRIES, S. TH. 2012: Corpus linguistics, theoretical linguistics, and cognitive/psycholinguistics: towards more and more fruitful exchanges. – J. Mukherjee ja M. Huber (toim.), *Corpus linguistics and variation in English: Theory and description*, 41–63. Amsterdam: Rodopi.
- GRIES, S. TH. – S. C. DESHORS 2014: Using regressions to explore deviations between corpus data and a standard/target: two suggestions. – *Corpora* 9(1), 109–136.
- GRIES, S. TH. – M. HILPERT 2008: The identification of stages in diachronic data: variability-based neighbor clustering. – *Corpora* 3(1), 59–81.
- GRIES, S. TH. – A. STEFANOWITSCH 2004: Extending collocation analysis: A corpusbased perspectives on 'alternations'. – *International Journal of Corpus Linguistics* 9(1), 97–129.
- GRIES, S. TH. – S. WULFF 2005: Do foreign language learners also have constructions? Evidence from priming, sorting and corpora. – F. J. Ruis de Mendoza Ibáñez (toim.), *Annual Review of Cognitive Linguistics* 3, 182–200. Amsterdam: John Benjamins.
- GRIES, S. TH. – S. WULFF 2009: Psycholinguistic and corpus-linguistic evidence for L2 constructions. – F. J. Ruis de Mendoza Ibáñez (toim.), *Annual Review of Cognitive Linguistics* 7, 163–186. Amsterdam: John Benjamins.
- HERLIN, I. 1998: *Suomen kun*. Helsinki: Suomalaisen Kirjallisuuden Seura.

- HERLIN, I. – L. KOTILAINEN 2004: External factors behind cross-linguistic similarities. – O. Fischer – M. Norde ja H. Perridon (toim.), *Up and Down the Cline. The Nature of Grammaticalization*. Amsterdam: John Benjamins, 263–279.
- HERLIN, I. – L. KOTILAINEN 2005: Itsenäistyyvä infinitiivi: 2. infinitiivin inessiivin kehitys kirjakielen aikana. – I. Herlin ja L. Visapää (toim.), *Elävä kielioppi: suomen infiniittisten rakenteiden dynamiikkaa*. Helsinki: Suomalaisen Kirjallisuuden Seura, 258–291.
- HOPPER, P. J. 1987: Emergent grammar. – *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 139–157.
- HYLAND, K. 2004: *Genre and Second Language Writing*. Ann Arbor (MI): The University of Michigan Press.
- IKOLA, O. 2000: *Pieni nykysuomen opas*. Turku: Turun yliopiston suomalaisen ja yleisen kielitieteen laitos.
- IKOLA, O. – U. PALOMÄKI – A.-K. KOITTO 1989: *Suomen murteiden lauseoppia ja tekstikielioppia*. SKST 511. Helsinki: SKS.
- ITKONEN, T. – S. MAAMIES 2007: *Uusi kieliopas*. Helsinki: Tammi.
- IVASKA, I. 2011: Lausetyyppien sekoittuminen edistyneessä oppijansuomessa – näkökulmana eksistentiaalilause. – *Lähivertailuja. Lähivördlusi* 21, 65–85. DOI: <http://dx.doi.org/10.5128/LV21.03>.
- IVASKA, I. 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>.
- IVASKA, I. 2014b: Mahdollisuuden ilmaiseminen S1-suomea ja edistynyttä S2-suomea erottavana piirteenä. – *Lähivertailuja. Lähivördlusi* 24, 47–80. DOI: <http://dx.doi.org/10.5128/LV24.02>.
- IVASKA, I. 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.
- IVASKA, I. 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, M. 2013: *Suomen eksistentiaalilause toisen kielen oppimisen polulla*. Jyväskylä Studies in Humanities 220. Jyväskylä: Jyväskylän yliopisto.
- KURKI, T. 2016: [Sähköpostiviesti]. Lähetetty 13.3.2016. Kirjoittajan hallussa.
- LARSEN-FREEMAN, D. 1997: Chaos/complexity and second language acquisition. – *Applied Linguistics* 18(2), 141–165.
- LARSEN-FREEMAN, D. – CAMERON, L. 2008: *Complex Systems and Applied Linguistics*. Oxford: Oxford University Press.
- LAURANTO, Y. 1997: *Ensi askeleita paikallissijojen käyttöön. Espanjankielisten suomenoppijoiden sisä- ja ulkopaikallissijat konseptuaalisen semantiikan näkökulmasta*. Kakkoskieli 2. Helsingin yliopiston suomen kielen laitos.
- LINDÉN, E. 1961: Temporaalirakenne suomen murteissa. – *Virittäjä* 65(2) 194–210.
- LINDÉN, E. 1971: Temporaalirakenne kirjoitetussa kielessä. – *Virittäjä* 75(1) 23–34.
- MARTIN, M. 2013. Sentences and clauses as complexity measures in second language writing: a segmentation experiment. – M. Järventausta ja M. Pantermöller (eds) *Finnische Sprache, Literatur und Kultur im deutschsprachigen Raum – Suomen kieli, kirjallisuus ja kulttuuri saksankielisellä alueella*. VSUA 85. Wiesbaden: Harrassowitz, 185–198.
- MARTIN, M. – S. MUSTONEN – N. REIMAN – M. SEILONEN 2010: On becoming an independent user. – I. Bartning, M. Martin ja I. Wedder (toim.), *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research*, 57–80. Eurosla: Monograph Series (1).

- MEUNIER, F. 2012: Formulaic language and language teaching. – *Annual Review of Applied Linguistics* 32(1), 111–129.
- MUSTONEN, S. 2015: *Käytössä kehittyvä kieli: Paikat ja tilat suomi toisena kielenä -oppijoiden teksteissä*. Jyväskylä Studies in Humanities 255. Jyväskylä: Jyväskylän yliopisto.
- NOORDMAN, L. – F. DE BLIJZER 2000. On the processing of clausal relations. Teoksessa E. Couper-Kuhlen & B. Kortmann (toim.), *Cause - Condition - Concession - Contrast: Cognitive and Discourse Perspectives*, s. 35–56. Berlin: Mouton de Gruyter.
- O'KEEFE, A. – M. MCCARTHY – R. CARTER 2007: *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- OLSHTAIN, E. – S. BLUM-KULKA 1985: Degree of approximation: Nonnative reactions to native speech act behavior. – S. M. Gass ja C. G. Madden (toim.), *Input in second language acquisition*, 303–325. Rowley (MA): Newbury House.
- PAQUOT, M., H. HASSELGÅRD – S. OKSEFJELL EBELING 2013: Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora. – S. Granger, G. Gilquin, ja F. Meunier (toim.), *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead. Proceedings of the First Learner Corpus Research Conference (LCR 2011)*. Louvain-la-Neuve: Presses Universitaires de Louvain, 377–387.
- R Core Team 2015: *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- SCHUTZ, N. 2013: How specific is English for Academic Purposes? A look at verbs in business, linguistics and medical research articles. – G. Andersen ja K. Bech (toim.), *English Corpus Linguistics: Variation in Time, Space and Genre*. Amsterdam: Rodopi, 237–257.
- SEILONEN, M. 2013: *Epäsuora henkilöön viittaaminen oppijansuomessa*. Jyväskylä Studies in Humanities 197. Jyväskylä: Jyväskylän yliopisto.
- SIITONEN, K. 2014: Learners' dilemma: an example of complexity in academic Finnish. The frequency and use of the E infinitive passive in L2 and L1 Finnish. – P. Lintunen, M. S. Peltola ja M.-L. Varila (toim.), *AFinLA-e Soveltavan kielitieteen tutkimuksia* 6, 134–148.
- SIITONEN K. – M. MIZUNO 2010: Suomen monitahoinen possessiivisuffiksi ja suomenoppija. – *Lähivertailuja. Lähivördlusi* 19, 136–159. DOI: <http://dx.doi.org/10.5128/LV19.09>.
- SIITONEN, K. – J. NIEMELÄ 2011: Mitä pitkittäistutkimus paljastaa edistyneiden suomenoppijoiden kielitaidosta? – *Lähivertailuja. Lähivördlusi* 21, 242–279. DOI: <http://dx.doi.org/10.5128/LV21.10>.
- SIIVELT, K. – S. MUSTONEN 2013: Lähdekielen vaikutus ja kielitaitotasot: paikallissijojen kehitys oppijansuomessa. – *Lähivertailuja. Lähivördlusi* 23, 341–370.
- SINCLAIR J. 1991: *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SWALES, J. 1990: *Genre Analysis: English in Academic and Research Setting*. Cambridge: Cambridge University Press.
- TOLVANEN, M. 2004: *Lauseenvastikkeet ei-syntyperäisten kielenpuhujien ylioppilasaineissa*. Pro gradu -tutkielma. Helsingin yliopiston suomen kielen laitos.
- TOMASELLO, M. 2003: *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.
- VISK = A. HAKULINEN – M. VILKUNA – R. KORHONEN – V. KOIVISTO – T. R. HEINONEN – I. ALHO 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkkoversio, <http://scripta.kotus.fi/visk> URN:ISBN:978-952-5446-35-7.