

Lehtinen, E., S. Aaltonen, M. Koskela, E. Nevasaari & M. Skog-Södersved (toim.) 2011. *AFinLA-e Soveltavan kielitieteen tutkimuksia 2011 / n:o 3*. 35–47.

Ilmari Ivaska ja Kirsti Siitonen

Turun yliopisto

Avainrakenneanalyysi: Tapa tutkia oppijankielen lauserakennetta korpusvetoisesti

Every yet-to-be-acquired function of language overlapping with something already acquired triggers a process of rearranging the functions. This paper discusses how to study the ways and order of the acquisition of clause structure and introduces a new corpus-driven way to do that. A method we call *Keyststructure Analysis* provides us a toolkit we can use – together with a syntactically annotated learner corpus – to capture structures essential in different stages of language acquisition. By the aforementioned means we study two syntactical structures in advanced Learner Finnish: i) existential sentence and ii) intransitive sentence similar to the existential sentence. We find out that during the 15-month observation period the use of the existential structure, when compared to Finnish as a first language, turns from quantitative overuse to underuse and the use of the studied intransitive structure from underuse to overuse, respectively. The change is similar in all the five subjects' subcorpora.

Keywords: syntax, corpus linguistics, Finnish as a second language, keyststructure analysis

1 Johdanto

Pohdimme tässä artikkelissa, miten kielen omaksumisreittejä voidaan tarkastella ja miten muotojen ja funktioiden välille syntyvien yhteyksien järjestymistä voidaan havainnoida pitkäaikaisaineistossa syntaktisesti annotoidun eli koodatun oppijankielen korpuksen avulla. Tarkastelemme näitä seikkoja syntaktisten rakenteiden tutkimuksessa ja ehdotamme yhdeksi vaihtoehdoksi korpusvetoista lähestymistapaa kunnioittavaa avainrakenneanalyysia. Avainrakenteella tarkoitamme jotakin sellaista rakennetta, jonka esiintymiä on oppijankielessä joko enemmän tai vähemmän kuin kohdekielessä tyypillisesti (ks. luku 3). Artikkelimme on kahtiajakoinen: ensinnäkin esittelemme metodologista linjaustamme ja toiseksi hahmottelemme näiden linjausten soveltuvuutta pilottitutkimuksen avulla.

Sovellamme avainrakenneanalyysiä prototyypin eksistentiaalilauseeseen (engl. *Existential sentence*, ks. Jespersen 1924: 154–156; termeistä *eksistentiaalilause* ja *eksistentiaalinen lause* ks. Ivaska 2010a: 68) ja muotopiirteiltään samanlaisen *olla*-verbin sisältävän intransitiivilauseen tarkasteluun. (ISK (848) kutsuu myös tätä rakennetta kopulalauseeksi. Kyseisen tulkinnan kritiikistä ks. esim. Ivaska 2009: 93.) Tutkimuskysymyksemme on, miten näitä lauserakenteita käytetään kirjoitetussa oppijansuomessa (S2) ja ensikieltään suomenkielisten kirjoittamassa suomessa (S1). Tutkimme, millaisia yhtäläisyyksiä ja eroavaisuuksia tutkittavien lauserakenteiden käytössä ja esiintymisfrekvensseissä on S2:n ja S1:n välillä. Rakenteet ovat morfologisilta ominaisuuksiltaan samanlaisia, mutta – prototyypisessä käytössä – ne poikkeavat sanajärjestyksensä ja semanttisen sisältönsä osalta toisistaan huomattavasti.

Tutkimuksemme on pitkittäistutkimus, eli tarkastelemme valittuja ilmiöitä 15 kuukauden tarkastelujakson ajalta. Tutkimus ja siinä käytettävä aineisto on osa Turun yliopistossa käynnissä olevaa Edistyneiden suomenoppijoiden korpus -tutkimushanketta (LAS2, ks. esim. Ivaska & Siitonen 2009; Jantunen & Piltonen 2009).

2 Toisen kielen syntaksin kehityksen kuvaaminen

Monissa toisen kielen omaksumista käsittelevissä tutkimuksissa on todettu, että omaksuminen seuraa useimmiten ennakoitavissa olevia reittejä. Nämä reitit ovat keskenään samankaltaisia, vaikka saavutettava kielitaito vaihtelee merkittävästi eri oppijoiden välillä. (Ks. esim. Goldschneider & DeKeyser 2005: 61–64; Pienemann 2005: 41.) Toisaalta omaksumisjärjestystä ja sitä kuvaavia reittejä pidetään usein puhtaasti kielen todellisten ilmaisujen ja rakenteiden esiintymisfrekvenssien määrittäminä (ks. esim. Tomasello 2003). Uutta kieltä omaksuessaan kielenoppija muodostaa jatkuvasti uusia muoto–funktio-

yhteyksiä. Kun uusi omaksuttava funktio vastaa joiltakin osin jotain aiemmin omaksuttua, alkaa funktioiden uudelleenjärjestyminen. Oppijan välikieli (engl. *Interlanguage*, ks. Selinker 1972; vrt. myös Selinker 1992, erityisesti s. 217–250 ja s. 259–264) pyrkii yksiselitteisyyteen, mutta samanaikaisesti omaksuttavat kielenpiirteet rikkovat muotojen ja funktioiden yhteyksiä ja synnyttävät tilanteita, joissa tarjolla olevien vaihtoehtojen välillä on vapaa variaatio. (Ellis 1985: 95–96.) Pyrimme tässä artikkelissa kuvaamaan tätä prosessia niin määrällisesti kuin laadullisestikin.

2.1 Lausetyypit ja tutkittavat lauserakenteet

Tutkimme tässä artikkelissa kahta suomen lauserakennetta, jotka koostuvat morfologisesti hyvin samankaltaisista lauseenjäsenistä: niissä on nominatiivissa tai partitiivissa oleva NP (lausekkeen subjektiudesta ks. esim. Helasvuo 1996: 348–349; Helasvuo & Huumo 2010: 191–192), predikaattina on *olla*-verbi ja niissä on inessiivisijainen lokatiivinen adverbiaali. Todellisuudessa adverbiaali voi toki olla kummassakin tarkasteltavassa lauserakenteessa useassa eri sijassa. Pilottitutkimuksen tarkastelun yksinkertaistamiseksi olemme kuitenkin päätyneet tarkastelemaan pelkästään inessiivisijaisia tapauksia kummastakin rakenteesta. Tarkasteltavia lauserakenteita voidaan pitää suomen kielessä yleisinä, sillä *olla*-verbi on suomen frekventein verbi (Saukkonen ym. 1979 s.v. *olla*) ja tutkimuskohteena olevassa tekstilajissa lähes 10 % lauseista on eksistentiaalisia (Ivaska 2010a: 73–74). Prototyypisessä käytössä lauseiden merkitykset erottaa toisistaan ainoastaan niiden sanajärjestyksen perusteella. Merkitykset voidaan abstrahoida muotoihin 1) 'jossakin on joku/jokin/jotakin' ja 2) 'joku/jokin on jossakin'. Esimerkki 1 kuvaa tarkasteltavaa eksistentiaalista lausetta ja esimerkki 2 tarkasteltavaa intransitiivilausetta. Kunkin esimerkin lopussa oleva koodi kertoo lauseen sijainnin LAS2-korpuksessa.

- (1) Sen takia paikkaa, *jossa on lehmät*, kutsua lehmäläksi. (las2-12tt01te07lo13)
- (2) Puhekielessä ja joissakin murteissa postpositio on kieliopillistunut pitkälle ja *se on muodossa -kaa*. (las2-1tt01te07lo13)

Osa tutkijoista pitää näiden kahden lauserakenteen välistä kahtiajakoa perustavanlaatuisena osana suomen kielen syntaksia (ks. esim. Hakanen 1972; Huumo & Perko 1993), kun taas toiset näkevät suomen lauseopin rakentuvan useista samanarvoisista lausetyypeistä, jotka voidaan erottaa toisistaan niiden morfo-syntaktisten piirteiden perusteella (ks. esim. Hakulinen & Karlsson 1979; ISK: 848–849). S2-alalla kiinnostus tällaisiin lausetyypeiksi kutsuttuihin lauserakenteisiin on useimmiten ollut luonteeltaan pedagogista, ja esimerkiksi Siitonen on todennut syntaktisten lausetyyppien tarjoavan oivallisen perustan lauseopin omaksumiseen ja omaksutun kielitaidon soveltamiseen kirjoit-

tetussa kielessä (Siitonen 1993: 118–119). Ivaska on tutkinut eksistentiaalisia lauseita ja eksistentiaalisuuden ilmaisemista S2:ssa sekä oppijansuomen syntaksin horisontaalista (synkroninen) ja vertikaalista (diakroninen) vaihtelua (Ivaska 2010a; Ivaska 2010b; termeistä horisontaalinen ja vertikaalinen ks. Ellis 1988: 168–171). Siitonen on esitellyt fenno-ugristikongressissa 2010, miten *olla*-verbillisten lauseiden määrä muuttuu informanteilla 7 kk:n tarkastelujakson aikana.

2.2 Lauserakenteen ja sen muutoksen kuvaaminen korpustutkimuksessa

Sinclair kuvaa ensikielen ja oppijankielen perustavanlaatuista eroa sillä, että ensikielisen kielenkäyttäjän järjestelmää ohjaa idiomaattisuus, kun taas oppijankielessä vallitsee lähtökohtaisesti vapaa valinta (Sinclair 1991: 109–110). Tämä ero voi usein olla määrällistä. Valintaa ohjaavat mekanismit ovat monisyisiä, ja esimerkiksi Jantunen luonnehtii kielen ilmaisuja kontekstuaalisiksi merkitysyksiköiksi, jotka koostuvat kontekstuaalisista (välittömän esiintymisympäristön luomista) valintapreferensseistä ja esiintymisehdoista (Jantunen 2004: 15, 29). Sähköinen LAS2-korpus voi valaista oppijansuomen syntaksia uudelta kannalta, sillä se mahdollistaa esimerkiksi syntaktisten rakenteiden esiintymistodennäköisyyksien vertailun esimerkiksi eri rakenteiden välillä ja suhteessa kohdekielen.

Housen soveltaa korpustutkimusta tarkastellessaan englantia toisena kielenä omaksuneiden verbijärjestelmää. Housen jakaa muutoksen vapaasta valinnasta kohti idiomaattista jakaumaa kolmeen vaiheeseen: 1) prototyypivaihe, jossa ei variaatiota, 2) vapaa variaatio eri allomorfien välillä ja 3) funktioiden tarkentuminen ja lähentyminen kohdekielen kaltaista käyttöä. (Housen 2002: 96–97.) Omaksunisen on monissa, eri lähtö- ja kohdekieliä tarkastelleissa tutkimuksissa huomattu etenevän vastaavanlaisissa vaiheissa (ks. esim. Dietrich ym. 1995; Giacalone 1992; Ivaska 2010a). Vaikkakin varsinaisten prototyypivaiheen ilmiöiden määrä vähenee oppijankielen kohdekielistymisen myötä, voi tätä esiintyä rakenteiden tasolla vielä edistyneidenkin oppijoiden kielessä (ks. esim. Ivaska 2010a: 81–83). Tämä vaiheiden valossa tehtävä tarkastelu on perusteltua tässäkin artikkelissa.

3 Avainrakenneanalyysi korpustutkimuksen metodina

Tutkimuksemme hyödyntää korpusvetoista tutkimusotetta (engl. *Corpus-Driven Approach*), eli emme näe korpusta pelkkänä työkaluna haluamamme tutkimuskysymyksen selvittämiseen ja aineiston erotteluun. Korpusvetoisessa tutkimuksessa teoreettinen viitekehys muotoutuu niiden tendenssien ympärille, joita korpuksen avulla voidaan

nostaa tarkastelun kohteeksi. (Ks. esim. Tognini-Bonelli 2001: 84–85.) Useimpien toista kieltä käsittelevien korpustutkimusten tapaan emme keskity niinkään laadulliseen epäidiomaattisuuteen, vaan pyrkimyksenä on tarkastella eri varianttien esiintymisympäristöjä ja määrällisiä trendejä, jotka voivat edelleen paljastaa vaikkapa yliyleistämistä (engl. *overgeneralization*, ks. esim. Richards 1971: 206–207) tai välttämistä (engl. *avoidance*, ks. esim. Davies 1991: 155). Näin ollen laadullinen epäidiomaattisuus on vain yksi tarkasteluista muuttujista. Kun tutkittava kieli lisäksi on hyvin edistyneiden oppijoiden tuottamaa, epäidiomaattisuus liittyy usein nimenomaan määrälliseen epäidiomaattisuuteen.

Kutsumme *avainrakenneanalyysiksi* sitä tapaa, jolla lähestymme tutkimuskysymystämme. Avainrakenne on jokin sellainen rakenne, joka esiintyy oppijankielessä huomattavasti kohdekieltä useammin tai harvemmin. Mitä suurempi tämä ero on, sitä todennäköisempää on, että rakenteen käyttö oppijankielessä eroaa kohdekielestä. Erot voivat olla niin laadullisia kuin määrällisiäkin, ja määrällinen ero voi myös paljastaa laadullisen eron. Avainrakenneanalyysi on luonteeltaan samankaltainen kuin jo kauan korpustutkimuksessa käytetty ja edelleen useassa tutkimuksessa olennainen avainsana-analyysi (ks. esim. Stubbs 2010).

On luontevaa ajatella, että juuri tällaiset avainrakenteet ovat syntaksin osalta oppijankieltä leimaavia tekijöitä. Kuten Meisel ym. moniulotteisessa mallissaan (engl. *Multi-Dimensional Model*) kuvaavat, kielen omaksuminen etenee useimmiten tietystä järjestyksessä, vaikka omaksumisen taso vaihtelee oppijoiden välillä (Meisel ym. 1981: 120–132). Pienemann on myöhemmin kehittänyt ajatusmallia edelleen prosessoitavuusteorian (engl. *Processability Theory*, ks. esim. Pienemann 2005). Toisaalta kielen käytöstä lähtevässä ajattelussa (engl. *Usage-Based Models*) lähestytään kielen oppimista frekvenssien, tilastollisten mallien ja ylipäänsä ympäristön antaman lingvistisen kokemuksen kautta (ks. esim. Kemmer & Barlow 2000). Avainrakenteen frekvensseissä tapahtuvat muutokset kuvaavat myös oppijankielen ja kohdekielen välisessä suhteessa tapahtuvia muutoksia – useimmiten omaksumista. Näin ollen avainrakenneanalyysin keskeisiä elementtejä onkin 1) löytää korpuksen avulla rakenteita, joita voidaan edellä mainitun perusteella pitää avainrakenteita, 2) tarkastella rakenteen esiintymisympäristöä ja sitä, miten se on samanlainen tai poikkeaa kohdekielen esiintymisympäristöstä, ja 3) tarkastella rakenteen käytössä ja mahdollisesti itse rakenteessa tapahtuvia pitkittäismuutoksia.

Tässä tutkimuksessa tarkastelemme avainrakenneanalyysin keinoin kahta lauserakennetta. Lauseilla on neljä yhteistä morfosyntaktista piirrettä: 1) subjektin kaltainen NP, 2) predikaattina toimiva *olla*-verbi, 3) inessiivisijainen adverbiali ja 4) ei predikaatiivia, joka ohjaisi tulkitsemaan lauseen kopulaksi. Käyttämämme muuttuja on verbin ja subjektin keskinäinen järjestys. Varsinainen haku tehtiin raakakoodatusta aineistosta käsin (ei siis koneellisen haun avulla), sillä LAS2-korpuksen käyttöliittymä ei vielä ole

käytettävissä ja saatavilla on vasta pilottitutkimus-aineistoa. Artikkelimme aineistontarkastelu pyrkii kuitenkin vastaamaan LAS2-korpuksen tulevaisuudessa mahdollistamaa hakua.

4 LAS2-korpus ja käyttämämme tutkimusaineisto

Tutkimusaineistomme on edistyneiden suomenoppijoiden kirjoittamia tenttivastauksia, ja se on kokonaisuudessaan osa LAS-korpusta. LAS2-korpuksen kaikki informantit ovat aineiston keräämishetkellä opiskelleet suomea jossain Turun yliopiston S2-oppijoille suunnatussa suomenkielisessä koulutusohjelmassa. (LAS2-korpuksen aineistosta tarkemmin ks. esim. Ivaska & Siitonen 2009; Jantunen & Piltonen 2009.) Kuten Granger toteaa, oppijankielen aineisto on harvoin täysin luonnollista. Tenttivastaukset ovat kuitenkin autenttisia kirjoitetun kielen yksikköjä siinä mielessä, että niitä ei ole kirjoitettu kielentutkimuksen aineistoksi. (Vrt. Granger 2002: 8.)

Aineistomme on viideltä edistyneeltä S2-oppijalta, ja sitä on yhteensä 20 002 sanetta. Vertailuaineistona (VA) on niin ikään LAS2-korpukseen kuuluvia S1-kirjoittajien tenttivastauksia, yhteensä 8 146 sanetta. Pyrimme tässä pilottitutkimuksessa tarkastelemaan mahdollisia avainrakenteita oppijoiden lähtökielestä riippumatta, joten mukana on typologisesti keskenään erilaisia ensikieliä. Ensikielien ovat seuraavat: englanti (LAS2-8, 4 406 sanetta), unkari (LAS2-10, 5 835 sanetta), islanti (LAS2-17, 2 738 sanetta), japani (LAS2-12, 4 336 sanetta) ja venäjä (LAS2-1, 2 687 sanetta). Tarkastelujakso on 15 kuukautta, ja tarkasteluajankohdat ovat 3, 6, 12 ja 15 kuukauden kohdalla. Mikäli informantilta on tallennettu tarkastelujaksolta vain yksi teksti, tarkasteluajankohta ei ole mukana kokonaistarkastelussa. 9 kuukauden kohdalla ei ole tarkasteluajankohtaa, sillä tarjolla oli vain hyvin niukasti aineistoa. Voidaan kuitenkin olettaa, että neljä valitsemaamme tarkasteluajankohtaa riittävät osoittamaan rakenteiden käytössä ilmeneviä muutostrendejä.

5 Tutkimustulokset

Käyttämämme aineisto sisältää yhteensä 81 hakuhehtomme täyttävää eksistentiaalista lausetta ja 41 intransitiivista lausetta. Vertailuaineisto sisältää puolestaan 23 hakuhehtoa täyttävää eksistentiaalista lausetta ja 9 intransitiivista lausetta. Eri osakorpusten tulosten vertailun helpottamiseksi olemme normalisoineet luvut tästä eteenpäin kuvaamaan esiintymisfrekvenssiä 100 000 saneen korpusnäytteessä (esiintymät / saneiden kokonaismäärä x 100 000). Normalisointi on yleinen malli tutkimusten ja aineistojen keskinäisen vertailtavuuden parantamiseksi (Biber ym. 1998: 263–264). Aineiston koko ja löydösten

määrä ovat niin pieniä, että niiden perusteella ei ole syytä tehdä kovinkaan vahvoja johtopäätöksiä itse tutkittavasta ilmiöstä. Normalisoiminen saattaa lisäksi vääristää merkittävästi käyttämästämme aineistosta saatavia tuloksia. Keskitymme tässä artikkelissa kuitenkin metodologiseen linjaukseen, minkä vuoksi käytämme huomattavasti suuremmille aineistoille paremmin sopivia työkaluja. Taulukko 1 kuvaa eksistentiaalisten lauseiden esiintymiä kussakin eri osakorpuksessa tarkastelujakson eri ajankohtina, ja taulukko 2 kuvaa vastaavia esiintymiä intransitiivilauseiden osalta.

TAULUKKO 1. Eksistentiaaliset lauseet jokaista 100 000 sanetta kohti aineiston eri osakorpuksissa. Informantti-sarake kertoo informantin, kk-sarakkeet tilanteen eri tarkasteluajankohtina ja VA-sarake vertailuaineiston esiintymämäärän.

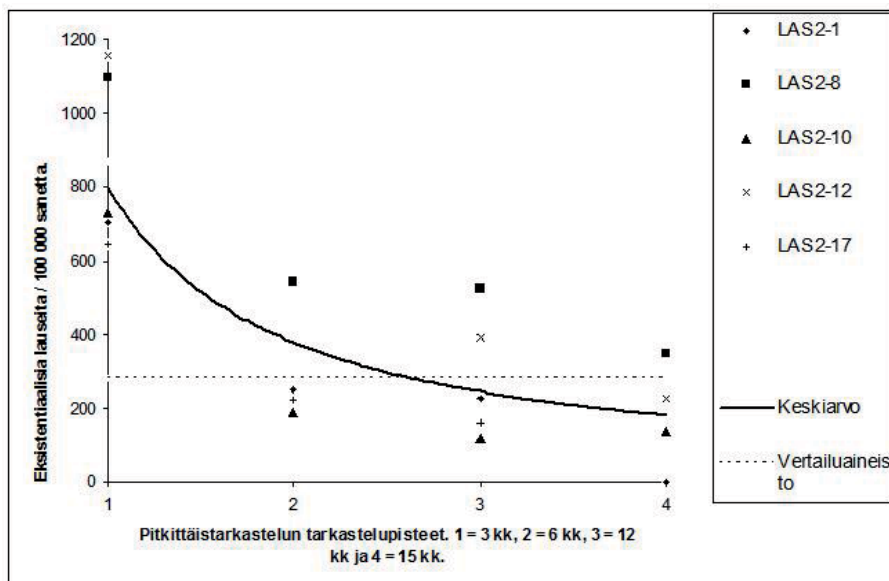
| Informantti | 3 kk | 6 kk | 12 kk | 15 kk | VA |
|------------------|------------|------------|------------|------------|------------|
| LAS2-1 | 702 | 251 | 227 | 0 | |
| LAS2-8 | 1099 | 544 | 525 | 347 | |
| LAS2-10 | 731 | 187 | 118 | 138 | |
| LAS2-12 | 1156 | | 390 | 228 | |
| LAS2-17 | 644 | 223 | 162 | | |
| Keskiarvo | 866 | 301 | 284 | 178 | 282 |

TAULUKKO 2. Tarkastelunalaiset intransitiivilauseet jokaista 100 000 sanetta kohti aineiston eri osakorpuksissa. Informantti-sarake kertoo informantin, kk-sarakkeet tilanteen eri tarkasteluajankohtina ja VA-sarake vertailuaineiston esiintymämäärän.

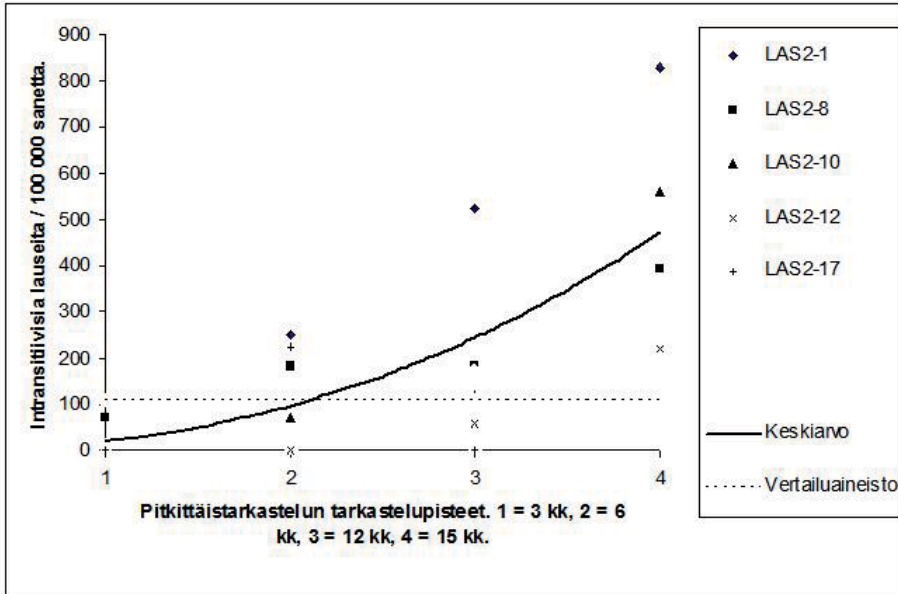
| Informantti | 3 kk | 6 kk | 12 kk | 15 kk | VA |
|------------------|-----------|------------|------------|------------|------------|
| LAS2-1 | 0 | 251 | 526 | 830 | |
| LAS2-8 | 69 | 183 | 181 | 394 | |
| LAS2-10 | 0 | 69 | 118 | 561 | |
| LAS2-12 | 0 | | 57 | 220 | |
| LAS2-17 | 0 | 223 | 0 | | |
| Keskiarvo | 14 | 145 | 176 | 501 | 110 |

Tarkastelujakson alussa tarkastelemiamme eksistentiaalisia lauseita on tutkimusaineistossa yli kolme kertaa niin paljon kuin vertailuaineistossa (866 / 100 000 sanetta vs. 282 / 100 000 sanetta). Tarkastelemamme intransitiivinen rakenne sitä vastoin esiintyy tarkastelujakson alussa tutkimusaineistossa paljon vertailuaineistoa harvemmin (14 / 100 000

sanetta vs. 110 / 100 000 sanetta). Eksistentiaalisten lauseiden määrä kuitenkin vähenee koko tarkastelujakson ajan, ja tarkastelujakson lopussa tutkimusaineistossa on eksistentiaalisia lauseita vähemmän kuin vertailuaineistossa (178 / 100 000 sanetta vs. 282 / 100 000 sanetta). Samanaikaisesti tarkasteltavien intransitiivisten lauseiden määrä kasvaa, ja tarkastelujakson lopussa niitä on tutkimusaineistossa lähes viisi kertaa niin paljon kuin vertailuaineistossa (501 / 100 000 sanetta vs. 110 / 100 000 sanetta). Kuvio 1 kuvaa havaitsemaamme muutosta eksistentiaalisten lauseiden osalta ja kuvio 2 kuvaa muutosta tarkastelemiemme intransitiivisten lauseiden osalta.



KUVIO 1. Eksistentiaalisten lauseiden esiintymät tarkastelujakson eri ajankohtina informanteittain. Käyrä kuvaa aineiston keskiarvoa ja sen muutosta ajan kuluessa.



KUVIO 2. Tarkasteltujen intransitiivisten lauseiden esiintymät tarkastelujakson eri ajankohtina informanteittain. Käyrä kuvaa aineiston keskiarvoa ja sen muutosta ajan kuluessa.

Kun aineistoa tarkastellaan laadullisen epäidiomaattisuuden (ks. esimerkit 3 ja 4) valossa, voidaan huomata, että tarkastelemiemme intransitiivisten lauseiden laadullinen epäidiomaattisuus tulee yleisemmäksi samalla, kun niiden suhteellinen määrä kasvaa ja eksistentiaalisten lauseiden suhteellinen määrä vähenee. Tarkastelujakson 6 ensimmäisen kuukauden aikana tarkastelemisemme intransitiivisissa lauseissa ei ole ainutakaan epäidiomaattista lausetta, 12 kuukauden tarkasteluajankohtana intransitiivisista lauseista 50 % on epäidiomaattisia ja 15 kuukauden jälkeen 43 % intransitiivisista lauseista on epäidiomaattisia (näitä osuuksia ei esitellä kuviossa 2). Lauserakenteiden semanttisen annin kannalta on mielenkiintoista, että epäidiomaattisuus vaikuttaa usein liittyvän kohdekielestä poikkeavaan erotteluun näiden kahden lauserakenteen välillä. Tyypillisessä epäidiomaattisessa tapauksessa lause esittää semanttisesti selvän väitteen jonkin asian eksistentiaalisuudesta, mutta lauseen pintarakenne on intransitiivinen. Esimerkit 3 ja 4 valottavat tätä ilmiötä.

- (3) Kontrastiivisessa analyysissä kiinnitetään huomiota seuraaviin seikkoihin: 1) mitä kahdessa kielessä on samaa; 2) mitä kahdessa kielessä on erilaista; 3) miten kahden kielen ero vaikuttaa toisen kielen oppimiseen ja omaksumiseen. Jos oppijan äidinkielessä on ilmiöitä, jotka ovat jollakin tavalla samanlaisia kuin toisessa kielessä, transfer voi tapahtua. **Enemmän transferia on fonetiikassa ja sanastossa*. (las2-1-tt01te08lo15)

- (4) Tekstin alussa löytyy OV-sanajärjestys, jossa on objektimainen subjekti ja yks. 3. persoonapäätteen kanssa verbi. Tätä kutsutaan tilalauseeksi, jossa ilmaistaan tilaa. **Samantyyppinen lause on tekstissä* 'Alkoi vielä sataakin'. Tekstin alussa myös käytetään lauseenvastiketta. (las2-12tt01te05lo03)

Molemmissa tapauksissa lauseen sanajärjestys on prototyyppinen intransitiiviselle lauseelle, mutta silti olemassaoloa kuvaava tulkinta on ainoa vaihtoehto. On totta, että nämä lauseet eivät ole yksiselitteisesti epäidiomaattisia, ja epäidiomaattisuus saattaa osin liittyä myös leksikaaliseen epäidiomaattisuuteen tai ympäröivään kontekstiin. Mikäli lauseet kuitenkin tulkitaan laadullisesti epäidiomaattisiksi, ovat ne osa eksistentiaalisten lauseiden ja niiden kaltaisten intransitiivisten lauseiden rajankäynnin kimppua. Tämän tulkinnan mukaan idiomaattiset lauseet olisivat jotakuinkin seuraavat (esimerkit 5 ja 6).

- (5) Fonetikassa ja sanastossa on enemmän transferia.
- (6) Tekstissä 'Alkoi vielä sataakin' on samantyyppinen lause.

Lisäksi esimerkissä 3 lauseen subjekti-NP:ksi jäsenyvä *transferia* on partitiivissa, mikä piirre on mahdollinen ainoastaan eksistentiaalisissa lauseissa (ISK: 849).

6 Huomioita ja keskustelua

Tulokset osoittavat selvästi havaittavan muutoksen kahdessa tarkastellussa lauserakenteessa. Tutkimus osoittaa hyvin avainrakenneanalyysin kyvyn nostaa korpusaineistosta esiin rakenteita, jotka ovat leimallisia oppijankielelle. Kuten todettua, oppijankieli ja sen muutokset seuraavat useimmiten jäljitettävissä olevaa kehityskulkua. Oppijankieltä tarkasteltaessa on mielestämme tärkeämpää paneutua sen dynaamisuuteen eli muutokseen enemmän kuin asiaintiloihin. Syntaktisesti annotoidulla korpuksella toteutettava avainrakenneanalyysi osana korpustutkimuksen metodologista kenttää tarjoaa oivallisen työkalun havaita tällaisia muutoksia ja pureutua muutosten luonteeseen. Tämän pilottiaineiston pieni koko ei kuitenkaan oikeuta vahvojen johtopäätösten tekemiseen itse tarkastelemastamme ilmiöstä. Seuraavia huomioita onkin syytä pitää enemmän valistuneina hypoteeseina ja suunnannäyttäjinä tuleviin tutkimuksiin kuin kattavina selvityksinä eksistentiaalisen rakenteen ja vastaavan intransitiivisen rakenteen käytöstä edistyneessä oppijansuomessa.

Vastataksemme tutkimuskysymyksiimme voimme verrata osoittamaamme muutosta Housenin (2002) kuvaamaan muutokseen englannin verbijärjestelmän omaksumisessa. Eksistentiaalinen rakenne on tarkastelujakson alussa lähes yksinomainen, kunnes tarkastelemamme intransitiivinen rakenne yleistyy ja kahden rakenteen välillä vallitsee vapaan valinnan kaltainen variaatio. Jakson aikana oppijoiden välikieli näyttää muuttuvan tarkasteltujen rakenteiden osalta. Tarkastelujakson lopussa funktiot eivät ole uudelleenjärjestyneet ainakaan kovin pitkälle ja tarkasteltujen lauserakenteiden funktionaalinen käyttö on edelleen kohdekielestä poikkeavaa niin määrällisesti kuin laadullisestikin. Tarkastelemamme intransitiivinen rakenne valtaa osan niistä funktioista, jotka olivat aiemmin eksistentiaalisen rakenteen toimialaa.

Esiintymisfrekvenssit vaihtelevat informanttien välillä runsaasti, mutta sama tendenssi toistuu kuitenkin kaikkien informanttien aineistoissa. Voidaankin olettaa, että näiden kahden rakenteen kohdekielen kaltainen omaksuminen ja käyttö ovat sidoksissa toisiinsa. Tendenssin toistuminen kaikkien informanttien kohdalla osoittaa myös, että ilmiö tuskin on sidoksissa oppijan lähtökieleen vaan nimenomaan suomeen toisena kielenä. Jos ilmiötä tarkastelee Meiselin ym. (1981) moniulotteisen mallin tai Pienemanin (2005) prosessoitavuusteorian kuvaamina vaiheina, vapaan variaation pitää edeltää rakenteiden kohdekielen mukaista käyttöä.

Avainrakenneanalyysi avaa uuden lähestymistavan oppijankielen syntaktisen kehityksen tarkastelemiseen. Syntaktisesti annotoidulla korpuksella on korpusvetoisen tutkimusotteen mukaisesti mahdollista etsiä rakenteita, jotka ovat avainasemassa oppijankielen kulloisessakin kehitysvaiheessa. Avainrakenteet ovat mielenkiintoisia ennen kaikkea silloin, kun tutkitaan syntaktisia muutoksia. Avainrakenteet sitoutuvat luonnollisiksi osiksi kielenomaksumisen jatkumoa ja avainrakenneanalyysi kykenee ainakin osittain kuvaamaan sitä, missä järjestyksessä rakenteita omaksutaan. Avainrakenneanalyysin suurimpia haasteita on annotoitujen korpusten infrastruktuurin luomisen lisäksi tarve muodostaa valistuneita hypoteeseja ja kyky muotoilla ne korpuksella kokeiltavaan muotoon, eli yhdistelmäksi leksikaalista, syntaktista ja morfologista metatietoa.

Kirjallisuus

- Biber, D., S. Conrad & R. Peppen 1998. *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.
- Davies, A. 1991. *Native Speaker in Applied Linguistics*. Edinburgh: Edinburgh University Press.
- Dietrich, R., W. Klein & C. Noyau 1995. *The Acquisition of Temporality in a Second Language*. Amsterdam: Benjamins.
- Ellis, R. 1985. *Understanding Second language Acquisition*. Oxford: Oxford University Press.
- Ellis, R. 1988. *Classroom Second Language Development*. London: Prentice Hall.

- Giacalone, R. A. 1992. Grammaticalization processes in the area of temporal and modal relations. *Studies in Second Language Acquisition*, 14, 297–322.
- Goldschneider, J. M. & R. M. DeKeyser 2005. Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language Learning* 55 no. *supp1*, 27–77.
- Granger, S. 2002. A Bird’s-eye view of learner corpus research. Teoksessa S. Granger, J. Hung & S. Petch-Tyson (toim.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins, 3–33.
- Hakanen, A. 1972. Normaali- ja eksistentiaalilause. *Sananjalka*, 14, 36–76.
- Hakulinen, A. & F. Karlsson 1979. *Nyky-suomen lauseoppi*. Suomalaisen Kirjallisuuden Seuran Toimituksia 350. Helsinki: SKS.
- Helasvuo, M.-L. 1996. Ollako vai eikö olla – eksistentiaalilauseen subjektin kohtalonkysymys. *Virittäjä*, 100, 340–356.
- Helasvuo, M.-L. & T. Huomo 2010. Mikä subjekti on? – *Virittäjä*, 114, 165–194.
- Housen, A. 2002. A corpus-based study of the L2-acquisition of the English verb system. Teoksessa S. Granger, J. Hung & S. Petch-Tyson (toim.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins, 77–118.
- Huomo, T. & J. Perko 1993. Eksistentiaalilause lokaalisuuden ilmaisijana. *Virittäjä*, 97, 380–399.
- ISK = Hakulinen, A., M. Viikuna, R. Korhonen, V. Koivisto, T. R. Heinonen & I. Alho, 2004. *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seuran Toimituksia 950. Helsinki: SKS.
- Ivaska, I. 2009. *Eksistentiaalilauseen ilmiösuhteiden tutkimus*. Turun yliopisto. Julkaisematon pro gradu -tutkielma.
- Ivaska, I. 2010a. Eksistentiaalilauseen subjektijäsen edistyneiden suomenoppijoiden kirjoituksessa. *Sananjalka*, 52, 67–88.
- Ivaska, I. 2010b. Mitä syntaktisesti koodattu korpus voi kertoa sanajärjestyksestä? Predikaatti ja subjekti edistyneiden suomenoppijoiden eksistentiaalisissa lauseissa. Teoksessa P. Eslon & K. Öim (toim.) *Korpusuuring ja metodid*. Tallinna ülikooli eesti keele ja kultuuri instituudi toimetised 12. Tallinna: Tallinna Ülikool, 37–58.
- Ivaska, I. & Siitonen, K. 2009. Syntaktisesti koodattu oppijankielen korpus: mahdollisuuksia ja ongelmia. Teoksessa P. Eslon & K. Öim (toim.) *Korpusuuringute metodoloogia ja märgendamise probleemid*. Tallinna Ülikooli Eesti Keele ja Kultuuri Instituudi toimetised 11. Tallinna: Tallinna Ülikool, 54–71.
- Jantunen, J. H. 2004. *Synonymia ja käännoosuomi: korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännoškielen leksikaalisiin erityispiirteisiin*. Joensuun yliopiston humanistisia julkaisuja. Joensuu: Joensuun yliopisto.
- Jantunen, J. H. & S. Pilttonen 2009. Oppijansuomen ja viron sähköiset tutkimusaineistot. *Virittäjä*, 113, 449–458.
- Jespersen, O. 1924. *The Philosophy of Grammar*. London: George Allen & Unwin Ltd.
- Meisel, J. M., Clahsen, H. & Pienemann, M. 1981. On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, 3, 109–135.
- Kemmer, S. & G. Barlow 2000. Introduction: A Usage-Based Conception of Language. Teoksessa G. Barlow & S. Kemmer (toim.) *Usage-Based Models of Language*. Stanford: CSLI, vii–xxviii.
- Pienemann, M. 2005. An Introduction to Processability Theory. Teoksessa M. Pienemann (toim.) *Cross-Linguistic Aspects of Processability Theory*. Amsterdam: Benjamins, 1–61.
- Richards, J. C. 1971. A non-contrastive approach to error analysis. *English Language Teaching Journal*, 25, 204–219.
- Saukkonen, P., M. Haipus, A. Niemikorpi, H. Sulkala 1979. *A frequency dictionary of Finnish*. Helsinki: WSOY.
- Selinker, L. 1972. Interlanguage. *International Review of Applied Linguistics*, 10, 209–231.

- Selinker, L. 1992. *Rediscovering Interlanguage*. London and New York: Longman.
- Siitonen, K. 1993. Peruslausetyypeistä johdettuja subjektittomia lauseita. Teoksessa J. Kalliokoski & K. Siitonen (toim.) *Suomeksi maailmalla. Kirjoituksia Suomen kielen ja kulttuurin opettamisesta*. Castrenianumin toimitteita 44. Helsinki: Helsingin yliopiston suomen kielen laitos, Helsingin yliopiston suomalais-ugrilainen laitos ja Suomalais-Ugrilainen Seura, 118–131.
- Siitonen, K. 2010. Mitä pitkittäistutkimus paljastaa edistyneiden oppijoiden kielitaidosta. Esitelmä fenno-ugristikongressissa 10.8.2010. Piliszsaba, Unkari.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, M. 2010. Three concepts of keywords. Teoksessa M. Bondi & M. Scott (toim.) *Keyness in Texts*. Amsterdam: Benjamins, 21–42.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: Benjamins.
- Tomasello, Michael 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.

Kiitämme Turun Yliopistosäätiötä LAS2-hankkeen saamasta tuesta.