

Oppimiskontekstin vaikutus oppijanpragmatiikkaan: astemääritteet leksikaalisina nallekarhuina¹

JARMO HARRI JANTUNEN

Jyväskylän yliopisto

Tiivistelmä. Artikkelissa käsitellään oppimisympäristön vaikutusta astemääritteiden käyttöön. Astemääritteet ovat jonkin ominaisuuden asteen suurta, kohtalaista tai vähäistä määrää ilmaisevia, pragmaattista merkitystä kantavia adverbeja (esim. *melko*, *hyvin*, *tosi*). Oppimisympäristön vaikutusta käsitellään artikkelissa korpusten avulla. Vaikka tällä hetkellä käytössä olevat oppijansuomen aineistot eivät ole täysin vertailukelpoisia keskenään esimerkiksi tekstien tehtävänantojen suhteen, voidaan niiden avulla tehdä alustavia havaintoja oppimiskontekstin vaikutuksesta ja edelleen hypoteeseja tulevia tutkimuksia varten. Oppimiskontekstilla tarkoitetaan tässä tutkimuksessa sitä ympäristöä, jossa kieltä opitaan: kielenoppimista kohdekulttuurin ulkopuolella (vieraana kielenä) tai kohdekulttuurissa (toisena kielenä). Aiemmassa tutkimuksessa on esitetty havaintoja, että oppimiskonteksti vaikuttaa oppimiseen, mutta myös vastakkaisia ajatuksia on tuotu esiin. Tutkimuksen menetelmällisenä lähtökohtana on avainsana-analyysi ja aineistona käytetään neljää korpusta (*Kansainvälinen oppijansuomen korpus*, *Yleisten kielitutkintojen korpus*, *Käännössuomen korpus* ja *Suomen kielen tekstikokoelma*), joiden hakutuloksia täydennetään internet-hauilla. Ensiksi mainittu aineisto sisältää vieraana

¹ Artikkelin perustuu AFinLA:n syysseminaarissa 2014 pitämäni esitelmään.

kielenä -aineistoa, toinen toisena kielenä -aineistoa ja kolme viimeistä kuvaavat kohdekielistä kielenkäyttöä. Tutkimustulokset eivät anna yksiselitteistä kuvaa astemääritteiden käytöstä: Ensimmäiset tulokset tukevat aiempia havaintoja siitä, että oppijan käyttävät astemääritteitä paljon, mutta erityisesti ne ovat suomea Suomessa opiskelevien suosiossa. Toisaalta kohdekulttuurissa opiskelevien astemääritekirjo osoittautuu kapeammaksi. Molemmilla ryhmillä on lisäksi käytössä ns. leksikaalisia nallekarhuja (engl *lexical teddy bears*), mutta nämä ovat osin erilaisia ryhmittäin. Astemääritteiden käyttö on sidoksissa syntagmaattisiin kombinaatioihin ja tilannekontekstiin, ja osoittautuu, että molempien ryhmien oppijat tuottavat jossain määrin fraseologisia tai rekistereihin liittyviä epätyypillisyyksiä.

Avainsanat: oppijankieli; oppimiskonteksti; korpustutkimus; avainsana-analyysi; n-grammit; astemääritteet

1. Johdanto

Nykyinen kielenoppimista ja oppijankieltä koskeva tutkimus on yhä useammin jollakin tavalla korpuksiin liittyvää, ja tutkimusalan kehityksessä sille on vakiintunut oma nimensäkin, *Learner Corpus Research*. Aineistoja kerätään ja kehitetään parhaillaan kovasti ja samalla joudutaan myös pohtimaan korpusmenetelmien soveltuvuutta oppijankielen tutkimukseen ja kehittämään uusia menetelmiä. Oppijankorpuksia hyödyntävässä tutkimuksessa valtavirtaa ovat olleet ensikielen vaikutuksen ja kielitaidon kehittymisen tarkastelu; se, mitä on tehty ennen korpuksia, on siis siirtynyt sittemmin korpustutkimukseen. Aineistoihin kerättävistä taustamuuttujista ovat tavallisimpia juuri oppijan ensikieli ja taitotaso (ks. esim. Jantunen & Pirkola 2015: 93), mikä mahdollistaa edellä mainittujen tutkimuskohteiden tarkastelun korpusmenetelmin. Se, mistä oppijankieliaineistojen materiaali kerätään, tuo luonnostaan tarjolle mahdollisuuden tarkastella myös oppimiskontekstin vaikutusta: Suomen rajojen ulkopuolelta kerättävät tekstit ovat pääsääntöisesti ns. vieraana kielenä -tekstejä, kotimaassa kerätyt puolestaan toisena kielenä

-tekstejä. Mukherjee ja Götz (tulossa 2015) painottavat kuitenkin, että oppimiskontekstin määrittely on ongelmallista, koska kontekstit eivät ole tarkkarajaisia eikä kielenoppija kuulu välttämättä yhteen kontekstiin vaan hänen oppimiskontekstinsa voi muuttua. Fyysisen kielenoppimisympäristön lisäksi kontekstiin voidaan laskea kuuluvaksi esimerkiksi opetusmenetelmät siinä missä opettajan äidinkielenkin. Aiempaan tutkimukseen nojaten Collentine ja Freed (2004: 157–161) päätyvät luokittelemaan oppimiskontekstit kolmeen pääasialliseen luokkaan: formaaliin luokkahuoneopetukseen (engl *formal language classroom*), kotimaassa annettuun kielikylpyopetukseen (*domestic immersion*) ja kohdekultuurissa suoritettuun opintojaksoon (*study abroad*). Näiden ulkopuolelle jää kuitenkin kokonaan kohdekultuurissa tapahtuva oppiminen, joka voi olla luokkahuoneopetuksessa tapahtuvaa oppimista, ns. naturalistista oppimista tai jotain näiden väliltä.

Vertaileva oppimisympäristöön kohdistuva tutkimus on alkanut saada jalansijaa hiljalleen 2000-luvulle tultaessa (Collentine & Freed 2004: 160–161). Nimenomaan korpustutkimuksen tavoitteeksi ehdottavat Mukherjee ja Götz (tulossa 2015) erityisesti *study abroad* -jaksojen merkityksen selvittämistä: jos korpuksista voidaan esimerkiksi poimia sellaiset oppijat, jotka ovat opiskelleet kohdekultuurissa, voidaan tarkastella, miten kohdekultuurissa opiskelu vaikuttaa kielitaitoon. Sen sijaan selvää vertailua puhtaasti kohdekultuurissa tapahtuvan (osittain tai kokonaan naturalistisen) oppimisen ja kohdekultuurin ulkopuolella tapahtuvan (lähinnä formaalin) oppimisen välillä eivät Mukherjee ja Götz alan kehitystä kuvaavassa, oppimiskontekstiin keskittyvässä korpuskatsauksessaan painota erikseen. Oppimiskontekstin vaikutuksen tutkimus ei olekaan saanut korpusanalyyseissä vielä kovin laajaa huomiota, mikä johtunee siitä, että yhteen korpukseen kootaan harvemmin aineistoa sekä toisena että vieraana kielenä -konteksteista; vertailtavat aineistot on siis kerättävä olemassa olevista korpuksista. Jonkin verran tutkimusta on kuitenkin jo tehty, ja esimerkiksi oppijansuomea on tarkasteltu kontekstivertailun näkökulmasta opinnäytetöissä (Virtanen 2011; Isohätälä 2014).

Koska oppijankorpuukset ovat tyypillisesti vain yhtä oppimiskontekstia edustavia, kohtaa tutkimus väistämättä ongelmia pohdittaessa, millaisia aineistoja verrataan keskenään ja miten vertailtavia ne ylipääntään ovat. Kun tavoitteena on pyrkiä mahdollisimman hyvään vertailtavuuteen, on esimerkiksi kirjoittajien ja tekstien taustamuuttujien oltava mahdollisimman samanlaisia. Tekstien osalta tämä koskee muun muassa tehtävänantoja, koska ne vaikuttavat suuresti erityisesti tuotoksen sanastoon. Edellä mainittuja tutkimuksia ja myös käsillä olevaa tutkimusta on pidettävä tästä syystä ainakin osittain pilottiluonteisena: vertailtavuus jää pakostakin hieman likimääräiseksi, vaikka erilaisilla taustamuuttujien määrittelyillä aineistoa valittaessa onkin pyritty mahdollisimman hyvään vertailtavuuteen.

Tässä tutkimuksessa verrataan oppijansuomen nykyaineistoja keskenään ja vertailuun otetaan mukaan myös natiiviaineistot. Analyysin kohteena ovat kielenoppijoiden tuottamat astemääritteet (luku 3), jotka paljastuvat avainsana-analyysin (5.1) näkökulmasta yhdeksi mahdollisesti oppimiskontekstin (2) vaikutusta heijastavaksi leksikaalis-pragmaattiseksi luokaksi. Astemääritteiden käyttöä käsitellään sekä esiintymistaajuuksien (5.2) että fraseologian (5.3) näkökulmasta.

2. Oppimiskontekstin merkitys

Oppimiskontekstin merkityksestä oppimistuloksiin ollaan SLA-alalla monta mieltä. Collentineen ja Freedin (2004: 153) mukaan alalla vallitsee kahtiajako jopa sen suhteen, pitäisikö konteksti ottaa ylipääntään huomioon tutkimuksessa: yhtäältä pidetään tärkeänä, että oppimisen prosesseja tutkitaan irrallaan taustamuuttujista, toisaalta intersektionaalisen näkökulman katsotaan tuottavan paremman kuvan oppimisesta ilmiönä. Oppimiskonteksteilla tarkoitan tässä tutkimuksessa siis kielenoppimista toisena kielenä- ja vieraana kielenä -ympäristöissä. Edellistä voi luonnehtia kärjistäen ja dikotomisesti implisiittiseksi, naturalistiseksi kohdekulttuurin sisällä tapahtuvaksi oppimiseksi (omaksumiseksi), jälkimmäistä puolestaan eksplisiittiseksi ja formaaliksi luokkahuoneoppimiseksi.

Näiden eroina on yleistäen pidetty muun muassa sitä, että naturalistisessa oppimisessa oppija pyrkii merkitykselliseen vuorovaikutukseen ja tavoitteena on sosiaalisen kompetenssin kehittyminen, kun taas formaali oppimiskonteksti tuottaa kielitiedon, rakenteet ja muodon hallitsevia analyttikkoja. Luokkahuoneoppijan ei ole myöskään katsottu käyttävän opittavaa kieltä juurikaan luokkahuoneen ulkopuolella, ja ainoa kohdekielinen kontakti saa olla opettaja, mutta toisena kielenä -oppijalla kielen katsotaan olevan kommunikoinnin väline arjessa monenlaisissa kielenkäyttötilanteissa. (Gass 1990; Batstone 2002; Collentine & Freed 2004; Ellis 2008: 109, 288, 302.) Kaikki eivät kuitenkaan halua tehdä eroa näiden kahden oppimiskontekstin välille, vaan sen sijaan korostetaan, että oppimisprosessit voivat olla samanlaisia ympäristöstä riippumatta, ja vaikka oppimiskontekstin voitaisiinkin katsoa vaikuttavan, riippuu pitkälti myös oppijan kielellisistä ja kognitiivista taidoista, millaisiksi hänen kommunikaatiotilanteensa rakentuvat näissä konteksteissa (Segalowitz & Freed 2004: 196; Ellis 2008: 288).

Oppijoiden tuotoksista on pyritty löytämään vastauksia siihen, vaikuttaako oppimiskonteksti jollakin tapaa oppimiseen; tulokset ovat osittain ristiriitaisia. Englannin artikkelisysteemin oppimista tarkastellut Ekiert (2004: 16, 18) päätyy toteamaan, että molemmissa konteksteissa oppimisen polut ovat samantapaiset eikä konteksti vaikuta tuotokseen. Samoin ajattelee oppimiskontekstien eroja teoreettisesti tarkasteleva Schinke-Llano (1990) korostaessaan, että oppimismenetelmät ovat ratkaisevia – eivät välttämättä oppimiskontekstit sinänsä. Sen sijaan esimerkiksi Gass (1987), Bardovi-Harlig ja Dörnyei (1998), Barron (2003), Collentine ja Freed (2004), Schauer (2009) ja Håkansson ja Norrby (2010) ovat kuitenkin löytäneet todisteita, että oppimiskontekstillä on merkitys oppimisessa ja että kontekstin vaikutus näkyy myös tuotoksessa. Pragmaattisten taitojen hallitsemista tarkastelleen Barronin (2003) laaja selvitys kohdekulttuurissa opiskelun vaikutuksesta paljastaa, että opiskelu kohdekulttuurissa on omiaan parantamaan nimenomaan pragmaattisia taitoja: pragmaattiset rutiinit ja esimerkiksi kieltäytymisen keinot muuttuvat enemmän kohdekulttuurin kaltaisiksi ja oleskeluajan

pituus kohdekulttuurissa vaikuttaa positiivisesti pragmaattisten taitojen kehittymiseen (2003: 238–247). Myös Bardovi-Harling ja Dörnyei (1998) toteavat, että toisena kielenä -opiskelijoiden pragmaattinen tietoisuus on parempi kuin vieraana kielenä verrokkiryhmän. Samoin Takahashi ja Beebe (1987) raportoivat ensimmäisen ryhmän hallitsevan paremmin kohdekultuurin pragmaattiset normit. Vaikka L1 heidänkin mukaansa aiheuttaa transferia molempien ryhmien oppijanpragmatiikkaan, se näyttää vaikuttavan enemmän vieraana kielenä -opiskelijoiden tuotokseen. Edelleen samanlaista toisena kielenä -opiskelijoiden parempaa pragmaattista, ja myös sanastollista, hallintaa ovat havainneet myös Schauer (2009) sekä Håkansson ja Norrby (2010), joskin jälkimmäiset myös toisaalta havaitsivat, että kieliopilliseen kompetenssiin ja sen kehittymisen polkuihin ei oppimisympäristöllä ole vaikutusta. Korpuspohjaisissa opinnäytetöissään oppijansuomesta sekä Virtanen (2011) että Isohätälä (2014) ovat havainneet oppimiskontekstin vaikuttavan jossain määrin tuotokseen: Virtasen verbien käyttöä koskeva tutkimus paljastaa vieraana kielenä -oppijoiden verbitaivutuksessa olevan enemmän ongelmia ja Isohätälä löysi puolestaan eroja sanojen käytössä – mielenkiintoisesti käsillä olevan tutkimuksen näkökulmasta eroja on esimerkiksi synonyymisten astemääritteiden valinnassa ja puhekielisten ilmausten määrässä sen mukaan, oliko aineisto toisena vai vieraana kielenä -aineistoa.

3. Astemääritteet

3.1. Astemääritteet pragmaattisina elementteinä

Astemääritteet (eli intensiteettiadverbit) ovat jonkin ominaisuuden asteen suurta, kohtalaista tai vähäistä määrää ilmaisevia adverbeja. Syntaktisesti ne ovat adjektiivien ja adverbien laajennuksia, tarkemmin määritteitä (VISK § 611, § 615, § 657), ja ne luokitellaan puoli-kieliopilliseksi sanoiksi, joilla ei ole suoraa tarkoitetta kielenulkoisessa maailmassa. Niille on yhteistä, että ne suhteuttavat jonkin adjektiivin ominaisuuden, tavan, ajan tai paikan adverbien sekä kvanttoriadverbien

ilmaiseman määrän johonkin puhujan pitämään tai yleisesti normaalin tai odotuksenmukaisena pidettyyn abstraktiin laatuun, tapaan tai määrään. Orpanan (1988: 153) mukaan kyse on ominaisuuden asteen tarkentamisesta. Luokalle on tyypillistä laaja synonyymikirjo (ks. Orpana 1988; VISK § 664; Jantunen 2004: 70–73), sillä astemääritteet ovat voimakkaasti tekstilaji- ja rekisterisidonnaisia ja samaa merkitsevät astemääritteet ovat erikoistuneet erilaisiin käyttökonteksteihin (vrt. *erittäin hyvä – tosi hyvä – sika hyvä – vitun hyvä*). Biberin (1988: 129–131) tulosten mukaan astemääritteiden käyttö on tavallista teksteissä, jotka osoittavat henkilökohtaista sitoutumista ja asenteita; tällaisia ovat hänen mukaansa muun muassa kasvokkais- ja puhelinkeskustelut ja henkilökohtaiset kirjeet. Lim ja Hong (2012: 134, 157) puolestaan painottavat, että tietyt astemääritteet ovat tyypillisiä teksteille, joissa asenteiden esittäminen on tavallista – toisia käytetään puolestaan neutraalimmissa teksteissä, kuten akateemisessa kirjoittamisessa (ks. myös Biber ym. 1999: 565). Myös muodoltaan ne ovat hyvin varioivia: osa leksikaalisuudesta määritteistä on muodostustavaltaan lähes opaakkeja (*lähes, järin*), osa läpinäkyviä adjektiivien tai substantiivien genetiivimuotoja (*mahdottoman, hirveän; ihmeen, helvetin*) ja osa puolestaan esimerkiksi nominatiivimuotoisia (*tosi, sika*). Myös komparatiivi- ja superlatiivimuotoja käytetään (*vähemmän kiltti, eniten kauhuissaan*). (VISK § 615, § 641; ks. myös Orpana 1988: 152–172.) Syntagmaattisesti astemääritteet ovat samanlaisia kuin muutkin sanat: toisilla luokan jäsenistä on jähmeähköjä, vähän variaatiota ilmentäviä kombinaatiosuhteita (esim. *tyystin* yleensä negatiivisessa ja *oikein* positiivisessa kontekstissa), toiset taas toimivat eräänlaisina yleismääritteinä, joiden kombinatoriset mahdollisuudet ovat vapaampia (*hyvin, todella, erittäin*) (ks. mm. Altenberg 1991; Jantunen 2004).

Pragmaattisesta näkökulmasta astemääritteet luonnehtivat, missä määrin käsiteltävänä oleva ominaisuus pätee kielenkäyttäjän mielestä. Jokin asia voi olla puhujan/kirjoittajan mielestä esimerkiksi *aika, melko* tai *suhteellisen järkevä*, jonkun toisen mielestä taas esimerkiksi *hyvin, erityisen* tai *sairaana järkevä*. Astemääritteillä ilmaistaan siis subjektiivisia

tulkinnallisia skaaloja (VISK § 666), emootioita (Orpana 1988: 156, 158–159; Klein 1998: 6), asennetta tai arviota (Orpana 1988: 165; Lorenz 1999: 24) tai sitä, miten asiaan tai sanottuun halutaan sitoutua (Paradis 1997: 10; Lorenz 1999: 24). Erityisesti tämä näkyy genetiivimuotoisissa määritteisä, jotka ilmaisevat selvästi myös muuta kuin määrää: *harvinaisen järkevä, riittävän usein* (Orpana 1988: 152–166; VISK § 664, § 675). Astemääritteet osoittavat siis kielenkäyttäjän osallisuutta: niiden käytöllä voidaan osoittaa, miten suuressa arvossa jotakin pidetään tai miten merkittäväksi piirteeksi tai määräksi jokin nähdään. Funktioltaan niiden käyttö on episteemisen modaalisuuden kaltaista. Osa astemääritteisistä voidaan laskea kuuluviksi varausten luokkaan, joille on tyypillistä epämääräistäminen (*suhteellisen nätti, vähän parempi*), osa taas voidaan luokitella intensifikaattoreiksi, jotka intensifioivat jotakin (*hirveen nätti*) (Stenström 1999), joskaan luokittelu ei ole päivänselvää (mm. puhekielinen *ihan* voi kuulua tilanteittain ja prosodian perusteella molempiin luokkiin). Edellä mainitun lisäksi astemääritteet toimivat myös ryhmään kuulumisen merkitsiminä: sukupolvilla on omat suosikkimääritteensä, ja uudet sukupolvet luovat ajalleen ominaisia määritteitä (Lorenz 1999: 25–26; ks. myös Lappalainen 2001: 80). Luokka onkin produktiivinen, muuttuva ja kirjava: se saa jatkuvasti uusia jäseniä (*sairaan hyvä, über kaunis*) ja niitä ketjutetaan (*ihan sairaan sika siisti; über ypöyksin*) olemassa olevien menettäessä ilmaisuvoimaansa; tilalle tarvitaan tuoreita ja intensiteetiltään esimerkiksi voimakkaamman tuntuisia sanoja (ks. Bolinger 1972: 18–19; VISK § 664.) Tämän lisäksi astemääritteiden käyttö voi olla myös kvantitatiivisesti ja kvalitatiivisesti sukupuolittunutta (Stenström 1999; Yaguchi ym. 2010).

3.2. Astemääritteet ja leksikaaliset nallekarhut kielenoppimisessa

Adverbien käytön hallinta on yksi kielenosaamisen mittari, sillä niiden kohdekielen mukaiseen käyttöön kiinnitetään huomiota muun muassa kielitestien arvioinneissa. Adverbiongelmat voidaan luokitella

tyypillisesti syntaksiin (sanajärjestykseen), epätyypillisiin frekvensseihin tai adverbien valintaan liittyviksi tapauksiksi. Näistä erityisesti kaksi viimeksi mainittua pätee astemääritteisiin. (Pérez-Paredes & Sándchez-Tornel 2014: 180.) Lorenz (1999: 26–27) korostaa, että astemääritteiden tehtävät – sitoutumisen ilmaiseminen ja ominaisuuden asteen ilmoittaminen – ovat kielenoppijoiden näkökulmasta oleellisia joskin konstikkaita asioita opittaviksi ja oppijat ovat myös itse tietoisia astemääritteiden käytön merkityksestä kielenkäytössä. Puolikieliopillisina sanoina ne aiheuttavatkin ongelmia jo pelkästään sen vuoksi, että niillä ei ole viittausuhdetta kielenulkoiseen maailmaan; yleensäkin käsitteelliset ja abstraktit ilmaukset ovat vasta itsenäisen kielitaidon tason tunnusmerkkejä, joskaan astemääritteet eivät liene kuitenkaan käsitteellisyyden tason hankalimmasta päästä, sillä niitä käytetään jo alkeistasoillakin. Toisaalta on huomattava, että myös niiden tekstilaji- ja rekisterisidonnaisuudet, tyylivärit ja muut edellä mainitut tilannekontekstuaaliset ominaisuudet tekevät käytön vaikeaksi.

Kielenoppimista ja oppijankieltä koskevassa tutkimuksessa onkin tehty havaintoja astemääritteiden käytön problemaattisuudesta; ongelmia on yhtä lailla kielitaustaltaan kuin kielitaidon tasoltaankin erilaisilla oppijoilla. Epätyypillisistä frekvensseistä eli niin sanotusta yli- tai alikäytöstä suhteessa äidinkielen “normiin” ovat raportoineen monet: esimerkiksi tuttuja, yleisten ja fraseologiaaltaan vähärajoitteisten täyteyden asteen määritteiden (*totally, completely*) ylikäytöstä ovat raportoineet Granger (1998) sekä Hinkel (2002; 2003) ja toisaalta taas joidenkin määritteiden alikäytöstä (esim. *highly*) Granger (1998) sekä Granger ja Rayson (1998). Havaintojen perusteella yksittäisten määritteiden käyttö voi siis vaihdella luokan sisällä oppijoiden suosiessa vain joitakin tarjolla olevista määritteistä. Vähäisestä astemäärittekirjosta onkin tehty huomioita: Pérez-Paredes ja Díez-Bedmar (2012) ovat havainneet nuorten kielenoppijoiden tyytyvän vain joihinkin tavallisimpiin määritteisiin (*very, so, really, too*), ja lähes samanlainen luettelo löytyy Sabaté i Dalmaun ja Curell i Gotorin (2007) anteeksipyyntöjen intensifioinnissa käytettyjen astemääritteiden luettelosta (*very, really, so, terribly*); Grangerin

(1998) aineiston mukaan *very* on näistä erittäin frekventti astemääräte oppijoiden tuotoksissa muiden jäädessä vähemmälle käytölle. Selityksenä valinnoille voi tarjota sekä tuttuutta että määritteiden fraseologista monikäyttöisyyttä, sillä frekventeimmät astemääritteet soveltuvat moniin koteksteihin ja konteksteihin. Samalla tällaiset havainnot tukevat ajatusta kielenoppijoiden tuotosten leksikaalisesta yksinkertaisuudesta, joka saattaa olla tulosta pelkistämisstrategiasta (Faerch & Kasper 1984: 48–49) tai kuulua oppijantuotoksen inherentteihin piirteisiin (Jantunen 2008).

Havaintoihin epätyypillisistä frekvensseistä liitetään etenkin nykyisin oppijankielen fraseologisuus; tämä koskee astemääritteiden käyttöä siinä missä muutakin sanastoa. Grangerin (1998) urauurtava ranskalaisten ja Lorenzin (1999) kattava saksalaisten englanninoppijoiden tutkimus kiinnittävät huomiota juuri oppijankielen leksikaalisiin kombinaatioihin, jotka osaltaan selittävät astemääritteiden natiivikielestä poikkeavia määriä. Syiksi epätyypillisiin kombinaatioihin on näissä tutkimuksissa esitetty muun muassa äidinkielen vaikutusta (Granger 1998: 151) ja sanojen yleistämistä kollokaatioihin, joissa niitä ei tyypillisesti käytetä, sekä oppijan pyrkimystä olla luova (Granger 1998: 150; Lorenz 1999: 217). Granger (1998: 147) korostaakin, että astemääritteiden kollokationaaliset suhteet asettavat kielenoppijoille suuria haasteita, koska astemääritteiden kollokaatiosuhteet voivat vaihdella rajoittuneista kollokaatiosuhteista (*bitterly cold*, vrt. *mikroskooppisen pieni*; Orpana 1988: 157) assosiaatiosuhteisiin, joissa valinta on melko vapaa (*completely different/new/free...*, vrt. *hyvin kaunis/paljon/hiljaa...*). Jollei opetus eivätkä oppi- ja sanakirjatkaan anna eksplisiittisesti tietoa siitä, mikä määräte – tai sana ylipäättäänkin – esiintyy minkäkin tyypisessä kontekstissa ja tekstilajissa tai rekisterissä ja miten astemääritteitä käytetään fraseologisesti, voi tuloksena olla helposti sinänsä kieliopillisen rakenteen näkökulmasta kelpo tuotos, mutta kuitenkin fraseologisesti epätyypillinen rakenne tai tilannekontekstiin sopimaton ilmaus (ks. Jantunen 2009; oppijankielen fraseologisuudesta laajemmin Nesselhauf 2005; Meunier & Granger 2008).

Oppijansuomesta astemääritteitä on aiemmin sivuttu muutamassa korpustutkimuksessa, mutta varsinainen niihin keskittyvä tutkimus odottaa vielä tekijää. Tarkastellessaan *paljon*-kvanttoriadverbin epätavallista yleisyyttä oppijoiden tuotoksissa Jantunen (2007: 76; ks. myös Kallioranta 2009) havaitsi, että suomenoppijat ovat yleistäneet sen käytön tyypillisestä astemääritekäytöstä, jossa siis sanaa käytetään komparoitujen adjektiivien ja adverbien määritteenä (*hän on paljon mukavampi*), myös tavalliseksi positiivimuotoisten sanojen astemääritteeksi muiden suurta määrää ilmaisevien määritteiden joukkoon (*Se on paljon 'hyvin kaunis, kuuluisa ja kallis festivalii.*). Toisessa tutkimuksessa (Jantunen 2009) tarkastellaan puolestaan *aivan*-adverbia nimenomaan natiivikielen uusuksen näkökulmasta lähtökohtana kielenoppijoiden tuottamat epätyypilliset fraseologiset kombinaatiot (kuten määrityssuhde *aivan paljon*) ja etsitään vaihtoehtoja muun muassa kyseisen *aivan*-adverbin opetukseen fraseologisena yksikkönä. Isohätälän (2014) työ puolestaan paljastaa, että *ihan* ja *liian* olisivat Suomessa suomea oppineiden suosimia ja *oikein*, *hyvin* ja *tosi* puolestaan ulkomailla suomea opiskelleiden suosimia astemääritteitä.

Kielenoppijoiden luontainen tekstin tuottamisen strategia on luottaa sanoihin, jotka tuntuvat helpoilta käyttää – ja jotka tämän vuoksi yliyleistyvät sellaisiin käyttötilanteisiin ja merkityksiin, joissa niitä ei tyypillisesti käytetä (Ringbom 2007: 72–73). Tyypillisesti nämä sanat ovat kielen neutraalia ja keskeistä sanastoa, joka on opittu esimerkiksi opintojen alkutaipaleella ja joka soveltuu moniin käyttöyhteyksiin. Esimerkiksi englannin astemääritteistä tällaisia ovat Hasselgrenin (1994: 255) ja Johanssonin (2008: 125–126) mukaan *very*, *much* ja *a lot*, mikä tukee yllä mainittuja havaintoja mm. *very*-adverbin yleisyydestä. Oppijat pelaavatkin leksikaalisissa valinnoissaan helposti varman päälle (Sabaté i Dalmau & Curell i Gotor 2007: 303; Granger 1998: 148) ja hyödyntävät Hasselgrenin (1994) metaforaa käyttäen niin sanottuja leksikaalisia nallekarhuja (*lexical teddy bears*) eli siis ilmauksia, jotka ovat tuttuja ja turvallisia ja joissa ei siten oleteta tehtävän virheitä. Tällainen strategia johtaa kuitenkin helposti sekä tyyllillisiin ja fraseologisiin outouksiin,

joskaan ei kuitenkaan aina varsinaisiin ymmärtämistä vaikeuttaviin kielivirheisiin. Isoissa tekstiaineistoissa huomataan tämän vuoksi myös kielenainesten epätyypillisiä frekvenssejä.

4. Aineistot ja menetelmät

4.1. Käytetyt korpuksot ja aineistot

Tutkimuksen aineistona on käytetty neljää suomalaista sähköistä kieliaineistoa sekä Google-hakukoneen antamia esiintymiä. Korpuksista kaksi on oppijansuomen aineistoja ja kaksi niin sanottuja natiivisuomen aineistoja. Oppimiskontekstin vaikutusta olen tutkinut Kansainvälisen oppijansuomen korpuksen (ICLFI) ja Yleisten kielitutkintojen korpuksen (YKI) avulla (korpuksista tarkemmin ks. Jantunen & Pirkola 2015). Näistä ensin mainittu on suomi vieraana kielenä -aineisto, jonka tekstit ovat peräisin suomea ulkomaisissa yliopistoissa opiskelevalta suomenoppijoilta. YKI-aineisto on puolestaan toisena kielenä -aineisto, sillä sen materiaali on peräisin Suomessa kielitutkinnossa käyneiltä henkilöiltä, jotka ovat pääasiassa oppineet suomea Suomessa. Molemmat oppijansuomen aineistot sisältävät sekä fiktiivisiksi teksteiksi että asiateksteiksi luettavia kirjoituksia (mm. kertomuksia, kirjeitä ja argumentoivia kirjoitelmia), mutta tekstien tehtävänannot eivät ole samanlaisia. Tehtävänannot vaikuttavat luonnollisesti jonkin verran sanastotutkimuksen tuloksiin mutta eivät kuitenkaan taitotasomäärittelyihin. On oletettavaa, että kirjoitustehtävistä kumpuava sanasto nousee esiin avainsana-analyysin tuloksissa, mutta toisaalta tehtävänannon selvästi motivoima sanasto on helppo karsia pois avainsanalistoista. Molemmat aineistot on annotoitu Connexorin fdg-jäsentimellä, joka antaa tulokseksi morfologisen ja syntaktisen annotaation. Annotointi on tehty puoliautomaattisesti, sillä epätyypillisten ja natiivikiielestä poikkeavien tuotosten vuoksi jäsentimen tuottama analyysi joudutaan tarkistamaan käsityönä (annotointiprosessista ks. Jantunen ym. 2014). Annotoitujen aineistojen lisäksi tutkimuksessa on käytetty hyväksi myös näiden aineistojen raakatekstiversioita avainsana-analyysin tekemisessä.

Koska tutkimuksen tarkoitus on tarkastella pelkästään oppimiskontekstin vaikutusta, eikä esim. taitotason vaikutusta, kummastakin edellä mainitusta aineistosta olen valinnut tekstit, jotka on sijoitettu arvioinnissa samalle taitotasolle. Taitotasoksi olen valinnut Eurooppalaisen viitekehyksen mukaisen kynnystason eli B1-tason, jota YKI-testissä vastaa taso 3 (asteikolla 1–6). Tälle tasolle arvioituja tekstejä on molemmissa korpuksissa runsaasti ja niiden kirjoittajien äidinkielet vaihtelevat. Tässä mielessä materiaalia voi pitää edustavana. Aineistojen koot ovat selvästi erilaiset, mutta koska avainsana-analyysi ottaa tilastollisessa kaavassaan huomioon aineistojen koot, ei aineistojen kokoerolla ole merkitystä tuloksiin. ICLFI-aineisto on kooltaan 102 000 ja YKI-aineisto 23 500 sanetta, tekstejä on vuorostaan 1946 ja 342; annotoiduissa aineistoissa on tunnisteita 876 000 ja 176 000.

Koska kummallekaan yllä mainitulle oppijansuomen aineistolle ei ole kerätty verrannollista natiiviaineistoa, olen käyttänyt saatavilla olevia muita korpuksia. Natiiviaineistona olen käyttänyt Käännössuomen korpuksen (KS-korpus, Mauranen 2000) ns. alkuperäissuomen osakorpuksista sekä Suomen kielen tekstikokoelmaa (SKTK). Ensin mainittu sisältää sekä fiktiivistä kirjallisuutta että asiatekstejä, jälkimmäisestä muodostin neljästä eri puolilla Suomea ilmestyvästä sanomalehdestä (*Karjalainen*, *Kaleva*, *Aamulehti* ja *Turun Sanomat*) osakorpuksen. Näitä aineistoja täydensin vielä Google-haun tuloksilla, koska korpusaineistot ovat yhtäältä 1990-luvun lopulla ilmestyneistä teksteistä koottuja, ja siten jo osittain hieman vanhoja, ja koska halusin täydentää tekstilajikirjoa mahdollisimman laajaksi. Internet-aineiston avulla mukaan saadaan hyvin erilaisia tekstejä tekstilajien, rekistereiden, tekstien fokuksen, ilmestymisajan ja editoinnin suhteen.

4.2. Käytetyt menetelmät ja työkalut

Annotoimattomista aineistoista tehty avainsanalista toimii tämän analyysin lähtökohtana. Avainsanat tarkoittavat ilmauksia, jotka kuvaavat tutkittavaa aineistoa leksikaalisesti ja ovat tilastollisesta näkökulmasta

merkitsevästi frekventimpiä tutkimusaineistossa kuin jossakin valitussa vertailuaineistossa. Tilastollinen avainsana-analyysi onkin yksi korpusvetoisen tutkimuksen menetelmistä: siinä aineiston annetaan nostaa esiin potentiaalisia tutkittavia yksiköitä. (Avainsanaisuudesta ja menetelmästä ks. Scott & Tribble 2006; Bondi & Scott 2010.) Menetelmää on käytetty hyväksi oppijankielen tutkimuksessa aiemminkin: Jantunen (2011) on tarkastellut ns. oppijankielen avainelementtejä annotoidun aineiston avulla; nämä jakautuivat tutkimuksessa avaintunnisteisiin (eli kieliopillisiin tageihin), sisältöavainsanoihin (eli sanoihin, jotka kuvaavat tekstien aiheita) ja oppijankielen avainsanoihin (eli sanoihin, jotka olisivat tyypillisiä nimenomaan oppijantuotokselle eivätkä esimerkiksi tehtävänannon motivoimia). Ivaska (2014; 2015) on puolestaan tehnyt tutkimusta oppijankielen avainrakenteista (esim. yksikön translatiivi <sg tra>; Ivaska 2014: 183–184) tarkoituksenaan selvittää, mitkä rakenteet ovat tyypillisiä oppijankielelle ja miten aiemmin opitut kielet vaikuttavat kieliopillisten rakenteiden käyttöön ja esiintymiseen.

Kun sekä YKI- että ICLFI-aineistoista on tehty sanalistat (Wordlists-työkalu) ja näitä on verrattu toisiinsa Keywords- eli avainsanatyökalun avulla (WordSmith Tools -ohjelmapaketti, ks. Scott 2008), saadaan tietää, mitkä sanat yli- ja aliedustuvat tutkimusaineistossa vertailuaineistoon verrattuna. Oppimiskontekstin näkökulmasta kyse on siitä, mitä ilmauksia Suomessa suomea opiskelevat käyttävät tilastollisesti enemmän (tai vähemmän) kuin ulkomailla suomea opiskelevat. Seuraavassa luvussa käydään aluksi läpi avainsana-analyysin tulokset, minkä jälkeen siirrytään analysoimaan avainsanoja yksityiskohtaisemmin.

5. Tulokset

5.1. Lähtökohtana avainsana-analyysi

Tutkimuksen lähtökohtana on YKI-aineistosta tehty avainsana-analyysi. Se tuottaa tulokseksi listan ilmauksia – tässä tapauksessa sananmuotoja – jotka esiintyvät huomattavasti useammin YKI-aineistossa kuin tutkimuksen tässä vaiheessa vertailuaineistona toimivassa ICLFI-aineistossa.

Analyysin tuloksesta on aluksi poistettu ilmaukset, jotka johtuvat aivan selvästi tehtävänannosta: Kirjoitustehtävissä on ohjeistettu kirjoittamaan mm. sähköpostiviestejä (esim. nimellä *Maija Solki*), arviointeja kurseista tai lehdestä ja esittelemään perusteluja. Tehtävän tekijä on usein ottanut sanastoa suoraan tehtävänannosta, jolloin avainsana-analyysi nostaa luonnollisesti esiin sanoja kuten *Maija, Solki, kurssi, lehti, perustelu*. Tällaisten suorien kopioiden lisäksi moni muukin avainsanalistan (taulukko 1) ilmaus motivoituu kyllä tehtävästä (esim. tervehdykset kirjeen alussa), mutta niille ei ole kuitenkaan välttämättä sanasanaista mallia tehtävänannossa, eikä ilman tutkimusta voida väittää, etteivät suomea ulkomailla opiskelijat käyttäisi samoja ilmauksia, joskaan eivät ehkä näin taajaan.

TAULUKKO 1. *YKI-aineiston tilastollisesti merkitsevimmät avainsanat*

Avainsanat (sanamuodot) merkitsevyyssjärjestyksessä
minä, moi, tosi , sinulle, sinä, nähdään, sinun, se, oli, hei, hyvä, hyvää, huomenna, kiitos, jos, voit, terve, anteeksi, terveisiä, kiva, lisää, kuuluu, minulle, soita, haluat, koska, tulla, mun, me, nyt, liian , pian, t., eilen, mak-saa, terveisin, ruoka, terveisiä, rikki, hinta, euroa, no, paljon, mitä, aika , olet, sun, hauska, toivon, moikka, parempi, tulen, en, pahoillani, mulle, mä

Merkille pantavaa avainsanojen joukossa (taulukko 1) on epäviralliseen rekisteriin tai puhekielisyyteen viittaavien sanojen tai muotojen määrä. Näitä ovat mm. persoonapronominien pikapuhemuodot *mä, mun, mulle, sun*, tervehdys- ja lopetusrutiinit *moi, moikka, terve, nähdään, terveisiä* jne. sekä tietyt adjektiivit ja adverbit, kuten *kiva, hauska, tosi* ja *aika*. On huomattava, että näiden käytössä on kysymys suuremmasta käyttötaajuudesta suhteessa ICLFI-aineistoon, ei luonnollisestikaan virheistä tai edes luultavasti rekisteripoikkeamistakaan. Pikemminkin on todennäköistä, että YKI-aineiston kirjoittajat ovat kohdekielisessä ympäristössä saaneet enemmän puhekielistä syötöstä kuin ICLFI-aineiston kirjoittajat. Tätä tukevat avainsanaluettelon puhutulle kielelle tyypillinen pronominien paljous sekä (deiktisten) ajankohdan adverbien runsas määrä

(*aina, eilen, ensi, kohta, nyt, pian*) tai vaikkapa *asia*-sanana puhekielinen synonyymi *juttu*.

Kuten edeltä käy ilmi, aiemmassa tutkimuskirjallisuudessa on raportoitu yhtäältä astemääritteiden epätyypillisestä käytöstä ja toisaalta siitä, että oppijat suosivat tiettyjä ilmauksia toisten kustannuksella. Seuraavassa tarkastelussa keskitytäänkin taulukon 1 avainsanojen *tosi, aika ja liian* motivoimana (*tosi* on erityisen frekventti ja kolmanneksi merkitsevin avainsana) koko astemääritteiden luokkaan tarkoituksena selvittää, miten yhtäältä suomea Suomessa ja toisaalta suomea ulkomailla opiskelevat käyttävät astemääritteitä. Olen hakenut aineistoista Concord-ohjelmalla kaikki astemääritteet hakusyntaksilla <premod adv>, joka tuottaa tulokseksi listauksen määritteenä toimivista adverbeista, kuten esimerkissä (1). Hakutuloksesta on lopuksi poistettu esiintymät, joissa ei ole astemääritettä (c, d).

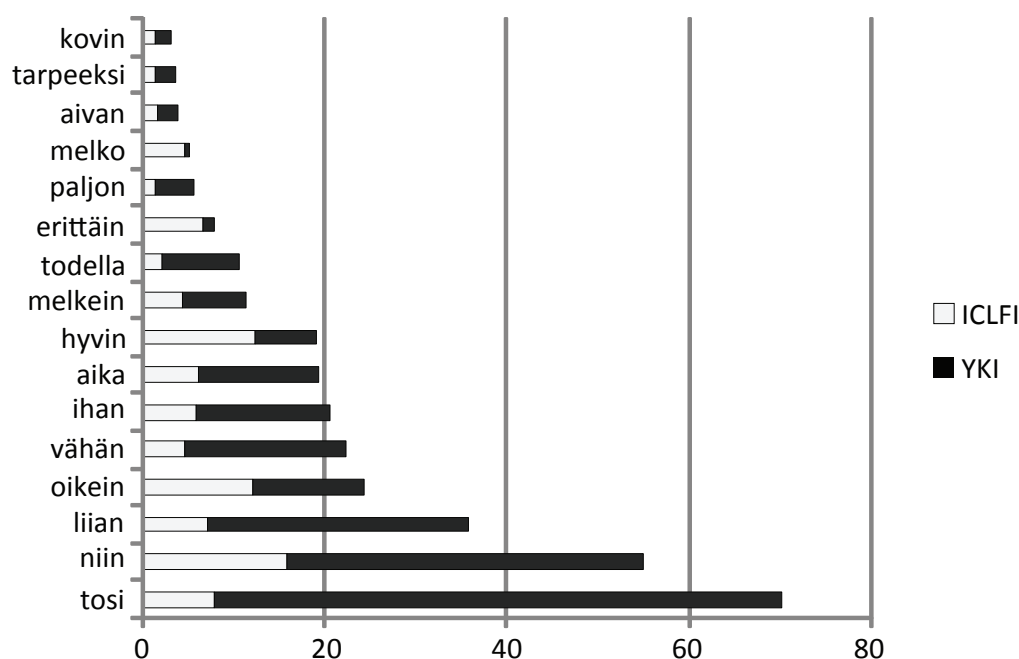
- (1) a. @ADVL ADV 4 hyvin hyvin @PREMOD ADV 5 lyhyet lyhyt @NH A SG NOM 6 2 2 @NH N
- b. @NH PRON NOM 2 erittäin erittäin @PREMOD ADV 3 hyvin hyvin @ADVL ADV 4 täällä tä
- c. @NH N SG PTV 46 1 lisää lisää @PREMOD ADV tehtävä tehtävä @NH N SG NOM 3 Talveits
- d. @MAIN V ACT IND PRES PL P3 4 vain vain @PREMOD ADV 5 2 2 @NH NUM CARD 6 eur

5.2. Astemääritteiden frekvenssit oppijankieliaineistoissa

Yki-aineisossa on astemääritteitä kaikkiaan 511, ICLFI-aineistossa 3453; suhteutettuna 10 000 saneeseen luvut ovat vastaavasti 221 ja 102. Aineistojen valossa näyttää siis siltä, että YKI-aineiston kirjoittajat ovat käyttäneet astemääritteitä yli kaksi kertaa niin paljon kuin ICLFI-korpuksen kirjoittajat. Eri astemääritteitä on YKI-aineistossa 19 ja ICLFI:ssä 39. Koska aineistot ovat erikokoiset, ei voi aivan suoraan vetää johtopäätöstä, että suomea ulkomailla opiskelevat käyttäisivät määritteitä variivammin: yksittäisetkin esiintymät ICLFI:ssä lisäävät kirjoa, ja tällaisia

yksittäisiä tapauksia voisi tulla vastaan YKI:stäkin, jos aineistoa kasvatettaisiin. Silti jonkinlaisesta variaation erilaisuudesta ja YKI-aineiston määritteiden yksinkertaisemmasta käytöstä voi lukumäärien ero kertoa. Tätä tukee myös frekventeimpien määritteiden osuus kaikista käytetyistä astemääritteistä: ICLFI:ssa kolme taajimmin käytettyä vie kaikista määritteistä 40 % osuuden, viisi puolestaan 55 %, kun taas YKI:ssä osuudet ovat vastaavasti 59 % ja 72 %. Lukujen valossa näyttää siis siltä, että YKI-aineiston kirjoittajat suosivat selvästi tiettyjä harvoja määritteitä.

Millaisia ovat sitten yksittäisten astemääritteiden frekvenssit aineistoissa? Kuviossa 1 on esitetty oppijankieliaineistojen astemääritteet suhteellisten frekvenssien mukaan. Kuviossa mainittujen lisäksi mm. *lähes, yhtä, juuri, kaikkein, vielä, riittävän, täysin* ja *yhä* (ICLFI) sekä *yhtä, noin ja kovasti* (YKI) ovat käytössä, mutta niiden suhteelliset määrät jäävät häviävän pieniksi (< 1).



KUVIO 1. Yleisimmät astemääritteet aineistoittain (x/10 000)

Kuviosta 1 käy ilmi, mitä määritteitä yleensä käytetään ja mitkä puolestaan ovat aineistoissa tavallisia. Ensinnäkin *oikein* on tasaisesti käytössä molempien aineistojen kesken, *hyvin*, *erittäin* ja *melko* ovat selvästi

ulkomailla opiskelevien suosiossa. *Niin, liian, ihan, aika, vähän, melkein, todella* ja etenkin *tosi* ovat suomea Suomessa opiskelevien käyttämiä määritteitä (vrt. Isohätälä 2014). Viimeksi mainituista *tosi, aika* ja *liian* ovat juuri niitä, jotka erottuivat taajaan käytetyiksi myös avainsana-analyysissä, mutta monet muutkin määritteet ovat siis nimenomaan YKI-aineistolle tyypillisempiä; tässä kohtaa avainsana-analyysin antama kuva siis tarkentuu. Aineistot poikkeavat täten selvästi sen suhteen, mitkä määritteet ovat tyypillisiä aineistoissa.

5.3. Oppijankielen astemääritefraseologiaa

Tarkastelemalla astemääritelaajennuksesta ja edussanasta koostuvia 2-grammeja (*n-gram*) eli lausekkeita saadaan selville, mitkä määrityysuhteet ovat tavallisimpia aineistoittain. Ne on lueteltu taulukossa 2. YKI-aineistossa on runsaasti 2-grammeja, joiden ensimmäisenä osana (eli määritteenä) on *tosi* tai *liian*, ICLFI-aineiston 2-grammeissa puolestaan nousee vahvasti esiin *oikein*-määrite (ks. myös Isohätälä 2014). 2-grammien jälkijäsenenä (eli lausekkeen edussanoina) on molemmissa tyypillisesti *hyvä* tai *paljon*, ja näistä jälkimmäinen korostuu erityisesti ICLFI-aineistossa. 15:n taajimman 2-grammin tarkastelusta paljastuu myös, että ICLFI-aineiston 2-grammeissa on enemmän variaatiota astemääritteiden käytössä. Tämä tukee edellä mainittua havaintoa YKI-aineiston suppeammasta astemääritekirjosta, sillä myös tavallisimpien astemääritegrammien astemääritekirjo näyttäisi siis olevan kapeampi.

Fraseologisesti 2-grammit eivät osoittaudu kovinkaan epätyypillisiksi. Natiivi-intuition perusteella yksikään ei tunnu mahdollomalta kollokaatiolta eikä kovin moni epätyypilliseltäkään. (Huomattakoon kuitenkin, että tässä yhteydessä 2-grammeja tarkastellaan irrallaan kotekstista, joten niiden tai astemääritteiden kollokoinnista saadaan verrattain kapea kuva eivätkä fraseologisesti epätyypilliset tapaukset tule välttämättä tässä esiin.)

TAULUKKO 2. *Frekventeimmät 2-grammit (astemääräite + edussana) lukumäärineen aineistoittain*

YKI-aineisto	ICLFI-aineisto
<i>tosi hyvä</i> 35	<i>niin paljon</i> 116
<i>tosi kiva</i> 28	<i>oikein paljon</i> 48
<i>niin paljon</i> 17	<i>liian paljon</i> 42
<i>liian paljon</i> 15	<i>oikein kaunis</i> 42
<i>ihan hyvä</i> 11	<i>erittäin paljon</i> 41
<i>oikein hyvä</i> 8	<i>tosi hyvä</i> 40
<i>liian pieni</i> 7	<i>oikein hyvin</i> 33
<i>liian kallis</i> 6	<i>hyvin paljon</i> 33
<i>liian vähän</i> 6	<i>oikein hyvä</i> 32
<i>niin hyvä</i> 6	<i>aika iso</i> 25
<i>tosi hyvin</i> 6	<i>tosi kiva</i> 24
<i>tosi paljon</i> 5	<i>melkein kaikki</i> 23
<i>tosi kaunis</i> 5	<i>liian väsynyt</i> 22
<i>liian pieni</i> 5	<i>oikein mukava</i> 21
<i>vähän tylsä</i> 5	<i>erittäin hyvä</i> 20

Intuitionkin perusteella jotkin 2-grammit tuntuvat kuitenkin potentiaalisesti hieman epätavallisilta. Tällaisia ovat *oikein*-määritteen sisältävät 2-grammit *<oikein paljon>* ja *<oikein kaunis>*. Natiiviaineistojenkin perusteella *<oikein paljon>* on harvinainen: koko KS-aineistossa niitä on vain 5 ja SKTK-aineistossa 13. Selvästi tavallisempia ovat *<hyvin/kovin paljon>*, joskaan vahventavan astemääritteen ja *paljon*-adverbien kolliointi ei ylipäätään ole kovin tavallista, vaan tyypillisempää on määrittää tätä adverbia deiktisellä *niin*-, vertailevalla *yhtä*- tai moderatiivisella *aika*-määritteellä (ks. myös taulukkoa 3). Kun vuorostaan kauneuden (*kaunis*) astetta halutaan vahventaa, tavallisimmat määritteet ovat natiiviaineistojen perusteella *niin*, *todella*, *erittäin*, *hyvin*, *erityisen*, *yhtä* ja *kovin*, eikä *oikein* ole siis tyypillisten joukossa.

Kun tarkastelussa fokusoidaan viimein 2-grammien (ks. taulukko 2) tavallisimpiin jälkijäseninä toimiviin sanoihin *paljon* ja *hyvä*,

tarkentuu kuva astemääritteiden käytöstä edelleen. Jotta voidaan tehdä vertailua myös natiiviaineistoon, on analyysiin otettu mukaan kolmesta natiiviaineistoista koottu yhdistetty luettelo tavallisimmista *paljon-* ja *hyvä-*sanojen määritteellisistä 2-grammeista. *Paljon-*adverbin osalta taulukosta 3 havaitsee ainakin kolme trendiä: Ensinnäkin YKI-aineiston avainsanaksi noussut *tosi* – joka on siis kyseisen aineiston frekventein ominaisuuden suurta astetta ilmaiseva, vahvistava astemäärite – poikkeaa käytöltään selvästi ICLFI-aineiston lisäksi myös natiiviaineistosta. Sen sijaan suomea ulkomailla opiskelevat käyttävät sitä lähestulkoon samassa suhteessa kuin natiivitkin kielenkäyttäjät. Toiseksi ominaisuuden kohtalaisen suurta määrää ilmaiseva *aika* on hieman harvemmin käytetty molemmissa S2-aineistossa natiiviaineistoon verrattuna, joskin YKI-aineisto noudattelee enemmän kohdekielistä käyttöä. Viimein suurta määrää ilmaiseva *oikein* paljastuu selvästi frekventisti käytetyksi määritteeksi nimenomaan ICLFI-aineistossa ja poikkeaa erityisesti natiiviaineistosta ja osin myös YKI-aineistosta, jossa se ei ole järjestyslistassa taajimmin käytettyjen joukossa. Vaikka *oikein* ei kokonaisvertailussa erottunutkaan erityisesti ICLFI-aineistossa käytetyksi määritteeksi, näyttää sen käyttö olevan runsasta juuri tässä aineistossa *paljon-*grammeja vertailtaessa. Mutta kun tarkastellaan vuorostaan *hyvä-*grammin määritteitä taulukosta 3, on kuva hieman toisenlainen, sillä nyt tässä käsiteltyjen astemääritteiden näkökulmasta oppijankieliaineistot ovatkin melko samankaltaisia (paitsi ehkä *ihan-* ja *erittäin-*määritteiden osalta) ja ne molemmat näyttävät poikkeavan natiivisuomesta: sekä *tosi* että *oikein* ovat molemmissa oppijanaineistoissa korkealla sijalla ja frekventisti käytettyjä. *Aika-*määritettä ei kuitenkaan tavata YKI-aineistosta. 2-grammien jälkijäseniä tarkasteltaessa saadaan siis aineistosta osittain jopa vastakkaista tietoa.

TAULUKKO 3. Määritteelliset paljon- ja hyvä-grammit yleisyysjärjestyksessä aineistoittain. (*NS-aineistosta Google-haut on tehty pelkästään paljon- ja hyvä-muodoista, muista NS-aineistoista nämä on haettu lemminoina)

< x paljon >			< x hyvä >		
YKI- aineisto	ICLFI- aineisto	Natiivi- suomi*	YKI- aineisto	ICLFI- aineisto	Natiivi- suomi*
<i>niin</i>	<i>niin</i>	<i>niin</i>	<i>tosi</i>	<i>tosi</i>	<i>niin</i>
<i>liian</i>	<i>oikein</i>	<i>yhtä</i>	<i>ihan</i>	<i>oikein</i>	<i>ihan</i>
<i>tosi</i>	<i>liian</i>	<i>aika</i>	<i>oikein</i>	<i>erittäin</i>	<i>erittäin</i>
<i>aika</i>	<i>erittäin</i>	<i>liian</i>	<i>niin</i>	<i>niin</i>	<i>yhtä</i>
<i>oikein</i>	<i>hyvin</i>	<i>kovin</i>	<i>todella</i>	<i>ihan</i>	<i>tosi</i>
<i>hyvin</i>	<i>aika</i>	<i>hyvin</i>	<i>erittäin</i>	<i>aika</i>	<i>aika</i>
<i>noin</i>	<i>kovin</i>	<i>tosi</i>	<i>kovin</i>	<i>todella</i>	<i>oikein</i>
	<i>tosi</i>	<i>erittäin</i>	<i>hyvin</i>	<i>melko</i>	<i>todella</i>
	<i>melko</i>	<i>melko</i>		<i>yhtä</i>	<i>kovin</i>
	<i>yhtä</i>	<i>varsin</i>		<i>hyvin</i>	<i>melko</i>
	<i>ihan</i>	<i>oikein</i>		<i>kovin</i>	<i>varsin</i>
		<i>ihan</i>			

6. Pohdintaa

Aineistojen vertailu osoittaa, että kielenoppijat käyttävät astemääritteitä paljon ja että niitä käyttävät enemmän Suomessa suomea oppineet kirjoittajat, mikä tukee aiempia havaintoja määritteiden käytöstä (mm. Granger 1998; Lorenz 1999; Hinkel 2002; 2003). Astemääritteiden käyttämisen onkin todettu olevan kielenoppijoiden tavallinen pragmaattinen keino: se on suosittu tapa pyrkiä vakuuttamaan lukija erilaisissa tekstilajeissa (Hinkel 2003: 1058). YKI-aineiston taajimmin käytetyt ja kokonaisfrekvenssiä selvästi nostavat astemääritteet ovat melko puhekielisiä, ja kirjakielisempiä käytetään vähemmän, toisin kuin ICLFI-aineistossa. Voineekin olettaa, että oppijat siirtävät kohdekulttuurisessa ympäristössä tapaamansa, sekä epävirallisen kirjoitetun kielen että

puhekielen, astemääritteiden käytön omaan kielenkäyttöön ja ylikorostavat tätä pragmaattista keinoa. Osittain kyse voi olla myös puhutun ja kirjoitetun kielen rekisterien sekoittumisesta (ks. Kalliokoski 1996; Gilquin & Paquot 2008; Jantunen 2008), sillä puhekielille tyypillisiä määritteitä tavataan aineistossa myös esimerkiksi tyyliltään kirjakielisimmistä mielipideteksteistä. Vastaavaa on havaittu myös englannin astemääritteiden käytöstä (Lorenz 1999: 215; Hinkel 2003). Tätä ei voi osoittaa kuitenkaan täysin pitäväksi ennen kuin vertailuaineistona on käytettävissä tekstilajeiltaan (ja tehtävänannoiltaan) samanlaisia natiivitekstejä. Jonkinasteinen puhekielisyyys teksteistä kuitenkin korostuu, mitä tukee erityisesti se, että taajimmin YKI-kirjoittajat käyttävät vahventavaa *tosi*-määritettä, joka on leimallisesti puhekielinen (VISK § 387, § 664), lisäksi puhekielessä taajaan käytettyjä ovat myös *aika* ja *ihan* (ks. VISK § 586, § 684, § 854, § 944). Kun tarkasteltiin *paljon*-adverbin saamia määritteitä, huomattiin, että erityisesti *tosi* poikkeaa natiivikäytöstä: se on suosituin valinta, kun YKI-aineiston kirjoittaja käyttää ominaisuuden suurta astetta merkitsevää astemääritettä. Sen sijaan kohtalaista astetta ilmaiseva *aika* noudattaa yleisyysjärjestyksensä puolesta natiivikäyttöä. Tässä on viitteitä siitä, että yliedustuminen ei ole kuitenkaan koko astemääriteluokkaa koskeva, vaan astemääritteet saattavat käyttäytyä yksilöllisesti. Näin ollen kuva pragmaattisten astemääritteiden ylikäytöstä ei ole yksiselitteinen. Vaikka Ellisin (2008: 288–290) mukaan kohdekielinen ympäristö on omiaan vahvistamaan pragmaattisia taitoja, astemääritteiden osalta selvää eroa ei voida tehdä oppimiskontekstin mukaan, sillä niiden käyttö varioi jopa määrityssuhdekohtaisesti.

Kun astemääritteitä tarkastelee vielä kokonaisuutena, näyttää siltä, että YKI-kirjoittajat keskittyvät käyttämään tiettyjä määritteitä, kun taas ulkomailla opiskelleilla kirjo on laajempi ja käyttö tasaisempaa. Tähän voi olla useita syitä. Ensinnäkin vaikutuksensa voi olla oppimismenetelmällä: osa YKI-testin suorittaneista on sellaisia, jotka ovat saaneet vähän tai eivät ehkä lainkaan formaalia opetusta vaan ovat oppineet suomen ns. luonnollisesti. Tällöin mm. sanaston käyttö mukautuu sen mukaiseksi, mitä kielenkäyttäjä kohtaa ympäristössään. Ulkomailla

opiskelleiden sanaston rikastuttajina toimivat lähestulkoon aina opettaja ja oppimateriaalit. Toiseksi on otettava huomioon tutkimusaineistojen tekstien kirjoitustilanne ja tekstintuottamistapa. YKI-aineiston tekstit on tuotettu koetilanteessa, jossa ei ole ollut käytössä tukimateriaalia, jolloin opiskelija (testiin osallistuja) saattaa ikään kuin pelata varman päälle ja käyttää hyvin tuntemiaan ja osaamiaan ilmauksia, leksikaalisia nallekarhuja. ICLFI:n teksteistä suurin osa on puolestaan tuotettu oppimistehtävinä, jolloin käytössä on voinut olla useita tukimateriaaleja eikä kielenopiskelijalla liene myöskään ollut samanlaista painetta turvautua tuttuihin ilmauksiin. Näiden lisäksi myöskään aineistojen osittain erilaisten tekstien ja tehtävänantojen vaikutusta ei voida sulkea pois. Huolimatta näistä taustamuuttujista ne eivät yksin selittäne tätä selvää eroa näiden tekstiaineistojen välillä.

Astemääritekohtaista valintaa heijastaa se, että esimerkiksi ominaisuuden suurta astetta merkitsevistä määritteistä toiset ovat erityisesti YKI- ja toiset ulkomaisten kirjoittajien suosiossa: kun *kovin* ja *oikein* tavataan lähes yhtä taajaan molemmista aineistoista, *erittäin-* ja *hyvin-*määritteitä suosivat ICLFI-kirjoittajat, *todella-* ja *tosi-*määritteitä YKI-kirjoittajat. Pelkkä taajuuden tarkastelu ei kuitenkaan riitä: kun määritteiden käyttöä tarkastellaan kvalitatiivisesti, huomataan, että vaikka frekvenssit tai yleisyys eivät juuri aineistoittain poikkeaisikaan, saattaa käyttö olla erilaista fraseologisesta näkökulmasta. Tästä kertoo esimerkiksi *oikein*, joka paljastui ICLFI-kirjoittajien taajimmin käyttämäksi vahventajaksi; tämä johtuu pitkälti fraseologisesti epätyypillisistä määrityysuhteista, joiden onkin oletettu olevan tavallisia kielenoppijoiden teksteissä (ks. mm. Hasselgren 1994; Granger 1998; Jantunen 2009). Toisaalta tämän tutkimuksen frekventeimmät 2-grammit eivät osoittaudu fraseologisesti kovinkaan epätyypillisiksi, kun 2-grammeja tarkastellaan kontekstista irrallaan. Epätavallisia rakenteita lienevät kuitenkin mm. alla olevan kuvion 2 kaltaiset ICLFI-kirjoittajien *oikein*-tapaukset: ne ovat kieliopin näkökulmasta hyvin rakennettuja määrite-edussanalausekkeitä, mutta fraseologisesti ongelmallisia mm. siksi, että vastoin kuin kuvio 2 antaisi olettaa, *oikein*-adverbia käytetään

natiivisuomessa tyypillisesti merkitykseltään (semanttiselta prosodi-
altaan) positiivisissa konteksteissa. Kielikontaktitilanteiden, mukaan
lukien esimerkiksi kääntämisen, onkin havaittu ainakin jossain määrin
vaikuttavan astemääritteiden fraseologiseen käyttöön (Jantunen 2004:
144–145, 214–215).

mista. Koripallo ja jalkapallo ovat mielestäni *oikein ikävä*. Kun kesällä lumi sulaa, niin käyn kaupungissa
lä. Menin retkelle ystäväni kanssa. Meidä oli *oikein paljon* Yövyimme teltassa ja laitoimme ruokaa itseemme
elitaitoni on keskimäinen, mutta en puhuu *oikein paljon* kieliä. Olen kinnostunut reggae-muusikista. Olen
kuljetti minua kotiin ja meni. Tiedän, että on *oiken vaikea* olla hymyhuulinen ja hyväntuulinen kokon päivän,
nglantia, vähän saksaa ja suomea. Ymmärrän *oikein vähän* venäjää. Olen kiinnostunut kauniista ilmasta ja.
siedä että joku voi olla viisampi. Turandot on *oikein ylpeä* ja sen seurauksena on sulhasen kuolema. Koska
ä mutta ei kukaan putoaa veteen. Minä olen *oikein ylpeä* jälkeensä. Me teemme tuli ja juomme kahvi

KUVIO 2. Oikein-grammeja ICLFI-aineistosta

7. Lopuksi

Sekä Ellis (2008: 290) että Mukherjee ja Götz (tulossa 2015) korosta-
vat, että oppimisympäristöjen jako yksioikoisesti luonnolliseen ympä-
ristöön ja luokkahuoneoppimiseen on liian karkea, koska yksikin
henkilö voi oppia kieltä erilaisissa konteksteissa. Ympäristön lisäksi
olisikin kiinnitettävä huomiota myös menetelmiin, joita käytetään ja
jotka näissä ympäristöissä tuovat parhaan mahdollisen oppimistulok-
sen. Tuotoksen tutkimuksessa tällainen jako on kuitenkin nähty tar-
peelliseksi, ja menetelmällisiä ratkaisuja on haettu korpustutkimuksen
keinoin. Kerätyillä materiaaleilla voidaan selvittää oppimisympäris-
tön vaikutusta kielenoppimiseen, mutta vaikka toiveena onkin saada
käyttöön oppimisympäristöjä ja oppijoita monipuolisesti edustavia,
taustamuuttujiltaan dokumentoituja korpuksia (Mukherjee & Götz,
tulossa 2015), on kuitenkin kiinnitettävä erityisesti huomiota myös
käytettävissä olevien aineistojen vertailtavuuteen (Jantunen & Pirkola
2015: 102). Tällä hetkellä esimerkiksi oppijansuomesta ei löydy tällai-
sia täysin taitotasojen, ensikielten, tekstilajien ja tehtävänantojen osalta

verrannollisia aineistoja. Esimerkiksi edellä kuvattu analyysi oppijansuomen astemääritteistä tuo kyllä esiin yhtäältä eroja, jotka johtunevat oppimisympäristöstä, ja toisaalta samankaltaisuuksia oppijanaineistojen kesken ja eroja suhteessa natiiviaineistoon (mikä puolestaan johtunevat kielenoppimisen prosesseista). Mutta tästä, kuten muistakaan tähän mennessä tehdyistä oppimiskontekstin ja tuotoksen suhdetta koskevista analyysistä, ei voida sulkea täysin pois aineistojen erilaisuuden merkitystä tuloksiin. Nykyisten S2-aineistojen suurin haaste onkin juuri vertailtavuuden ongelma, ja sama tilanne koskee osittain oppijanaineistojen vertailua natiiviaineistoihin (vrt. kuitenkin Ivaska 2014; 2015).

Sanasto, ja laajemmin fraseologia, on hyvin rekisteri- ja tekstilajisidoksista, eivätkä astemääritteet tee tässä poikkeusta. Pragmaattisina, sitoutumista ja asennetta ilmaisevina elementteinä sekä tyyllilaji- ja rekisteri-indikaattoreina niiden oikeanlainen kvalitatiivinen ja kvantitatiivinen käyttö osoittaa kielenkäytön hienovaraisten vivahteiden hallintaa tai sen puutetta, ja siksi niiden tarkastelu on tähdellistä – samoin kuin olisi laajemminkin oppijansuomen sanastopiirteiden. Menetelmien – ja aineistojenkin – kehittämisen vuoksi vertailuja ja tutkimusta kannattaa tehdä jo nykyisilläkin aineistoilla, kunhan pidetään mielessä analyysien reunaehdot.

Lähteet

- Altenberg, Bengt 1991. Amplifier collocations in spoken English. – Stig Johansson, Anna-Brita Stenström (Eds.). *English Computer Corpora. Selected Papers and Research Guide*. Berlin: Mouton de Gruyter, 127–147.
- Bardovi-Harlig, Kathleen, Zoltán Dörnyei 1998. Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. – *Tesol Quarterly* 32 (2), 233–259. <http://dx.doi.org/10.2307/3587583>
- Barron, Anne 2003. *Acquisition in Interlanguage Pragmatics: Learning How to Do Things with Words in a Study Abroad Context*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/pbns.108>

- Batstone, Rob 2002. Contexts of engagement: A discourse perspective on 'intake' and 'pushed output'. – *System* 30 (1), 1–14. [http://dx.doi.org/10.1016/S0346-251X\(01\)00055-0](http://dx.doi.org/10.1016/S0346-251X(01)00055-0)
- Biber, Douglas 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511621024>
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson.
- Bolinger, Dwight 1972. *Degree Words*. The Hague: Mouton. <http://dx.doi.org/10.1515/9783110877786>
- Bondi, Marina, Mike Scott (Eds.) 2010. *Keyness in Texts*. SCL 41. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/scl.41>
- Collentine, Joseph, Barbara F. Freed 2004. Learning context and its effects on second language acquisition. Introduction. – *Studies in Second Language Acquisition* 26 (2), 153–171. <http://dx.doi.org/10.1017/S0272263104262015>
- Ekiert, Monika 2004. Acquisition of the English article system by speakers of Polish in ESL and EFL settings. – *Columbia University Working Papers in TESOL & Applied Linguistics* 4 (1), 1–23.
- Ellis, Rod 2008. *The Study of Second Language Acquisition*. 2nd edition. Oxford: Oxford University Press.
- Faerch, Claus, Gabriele Kasper 1984. Two ways of defining communication strategies. – *Language Learning* 34 (1), 45–63. <http://dx.doi.org/10.1111/j.1467-1770.1984.tb00995.x>
- Gass, Susan M. 1987. The resolution of conflicts among competing systems: A bidirectional perspective. – *Applied Psycholinguistics* 8 (4), 329–350. <http://dx.doi.org/10.1017/S0142716400000369>
- Gass, Susan M. 1990. Second and foreign language learning: Same, different or none of the above? – Bill Van Patten, James F. Lee (Eds.). *Second Language Acquisition, Foreign Language Learning*. Clevedon: Multilingual Matters, 34–44.
- Gilquin, Gaëtanelle, Magali Paquot 2008. Too chatty. Learner academic writing and register variation. – *English Text Construction* 1 (1), 41–61. <http://dx.doi.org/10.1075/etc.1.1.05gil>
- Granger, Sylviane 1998. Prefabricated patterns in advanced EFL writing: Collocations and formulae. – Anthony P. Cowie (Ed.). *Phraseology: Theory, Analysis and Applications*. Oxford: Oxford University Press, 145–160.
- Granger, Sylviane, Paul Rayson 1998. Automatic profiling of learner texts. – Sylviane Granger (Ed.). *Learner English on Computer*. London: Longman, 119–131.

- Hasselgren, Angela 1994. Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. – *International Journal of Applied Linguistics* 4 (2), 237–260. <http://dx.doi.org/10.1111/j.1473-4192.1994.tb00065.x>
- Hinkel, Eli 2002. *Second Language Writers' Text. Linguistic and Rhetorical Features*. Mahwah NJ: Erlbaum.
- Hinkel, Eli 2003. Adverbial markers and tone in L1 and L2 students' writing. – *Journal of Pragmatics* 35 (7), 1049–1068. [http://dx.doi.org/10.1016/S0378-2166\(02\)00133-9](http://dx.doi.org/10.1016/S0378-2166(02)00133-9)
- Håkansson, Gisela, Catrin Norrby 2010. Environmental influence on language acquisition. Comparing second and foreign language acquisition of Swedish. – *Language Learning* 60 (3), 628–650. <http://dx.doi.org/10.1111/j.1467-9922.2010.00569.x>
- Isohätälä, Jaana 2014. Kielenoppimisympäristön vaikutus oppijansuomen sanastoon. Julkaisematon pro gradu -tutkielma. Oulun yliopisto, suomen kieli. <http://jultika.oulu.fi/Record/nbnfioulu-201405291585> (12.2.2015).
- Ivaska, Ilmari 2014. Edistyneen oppijansuomen avainrakenteita. Korpusnäkökulma kahden kielimuodon tyypillisiin rakenteellisiin eroihin [‘Key structures in advanced learner Finnish: Corpus approach towards structural differences between two language forms’]. – *Virittäjä* 118 (2), 161–193.
- Ivaska, Ilmari 2015. Tracing crosslinguistic influences in structural sequences: What does key structure analysis have to offer? – *Bergen Language and Linguistics Studies 6: Learner Corpus Research: LCR2013 Conference Proceedings*, 23–44.
- Jantunen, Jarmo Harri 2004. Synonymia ja käännessuomi: Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännessuomen leksikaalisiiin erityispiirteisiin [‘Synonymity and Translated Finnish. A Corpus-based View of Contextuality of Synonymous Expressions and Lexical Features Specific to Translated language’]. Joensuun yliopiston humanistisia julkaisuja 35. Joensuu: Joensuun yliopisto. http://epublications.uef.fi/pub/urn_isbn_952-458-479-4/urn_isbn_952-458-479-4.pdf (12.2.2015).
- Jantunen, Jarmo Harri 2007. Oppijansuomen piirteitä korpusvetoisesti [‘Corpus-driven analysis of learner Finnish’]. – Pirkko Muikku-Werner, Ossi Kokko, Hannu Remes (Toim.). *Virsu* 3. Suomalais-ugrilaisia kohdekieliä ja kontakteja. Joensuu: Joensuun yliopisto, 69–83.
- Jantunen, Jarmo Harri 2008. Haasteita oppijankielen korpusanalyysille: oppijankielen universaalit. – Pille Eslon (Toim.). *Öppijakeele analüüs: võimalused*,

- probleemid, vajadused. Tallinna ülikooli eesti filoloogia osakonna toimetised 10. Tallinn: TLÜ kirjastus, 67–92.
- Jantunen, Jarmo 2009. “Minulla on aivan paljon rahaa” – Fraseologiset yksiköt suomenkielen opetuksessa [‘I have really lots of money’ – Phraseological units in the teaching of Finnish]. – *Virittäjä* 113, 356–381.
- Jantunen, Jarmo Harri 2011. Avainsana-analyysi annotoidun oppijankieliaineiston tutkimisessa: Alustavia havaintoja [‘Keyword analysis and annotated learner data: Preliminary findings’]. – *AFinLA-e. Soveltavan kielitieteen tutkimuksia*, 48–61. <http://ojs.tsv.fi/index.php/afinla/article/view/4456> (2.2.2015).
- Jantunen, Jarmo Harri, Sisko Brunni, Liisa-Maria Lehto, Valteri Airaksinen 2014. Oppijankieliaineistojen annotointi: esimerkkinä ICLFI:n annotoinnin prosessit, ongelmat ja ratkaisut. – *AFinLA-e. Soveltavan kielitieteen tutkimuksia* 7, 60–80. <http://ojs.tsv.fi/index.php/afinla/article/view/48160> (3.9.2015).
- Jantunen, Jarmo Harri, Silja Pirkola 2015. Oppijansuomen sähköiset tutkimusaineistot: nykytilanne. – *Virittäjä* 119 (1), 89–104.
- Johansson, Stig 2008. *Contrastive Analysis and Learner Language: A Corpus-Based Approach*. Oslo: University of Oslo. http://www.hf.uio.no/ilos/forskning/grupper/Corpus_Linguistics_Group/papers/contrastive-analysis-and-learner-language_learner-language-part.pdf (15.1.2015).
- Kalliokoski, Jyrki 1996. Puhe, kirjoitus, tekstilajin normit ja toisella kielellä kirjoittaminen. – Helena Ruuska, Sanna-Marja Tuomi (Toim.). *Moneja baareja tiellä toimivaan kaksikielisyyteen*. Helsinki: Äidinkielen opettajain liitto, 107–119.
- Kallioranta, Otto 2009. *Paljon*-adverbin kollokointi oppijansuomessa. Korpusvetoinen tutkimus. Julkaisematon pro gradu -tutkielma. Oulun yliopisto, suomen kieli.
- Klein, Henny 1998. *Adverbs of Degree in Dutch and Related Languages*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/la.21>
- Lappalainen, Hanna 2001. Sosiolingvistinen katsaus suomalaisnuorten nykypuhekieleen ja sen tutkimukseen. – *Virittäjä* 105 (1), 74–101.
- Lim, Ni-Eng, Huaqing Hong 2012. Intensifiers as stance markers. A corpus study on genre variation in Mandarin Chinese. – *Chinese Language and Discourse* 3 (2), 129–166. <http://dx.doi.org/10.1075/cld.3.2.01lim>
- Lorenz, Gunter R. 1999. *Adjective Intensification: Learners Versus Native Speakers. A Corpus Study of Argumentative Writing*. *Language and Computers: Studies in Practical Linguistics* 27. Amsterdam: Rodopi.

- Mauranen, Anna 2000. Strange strings in translated language: A study on corpora. – Maeve Olohan (Eds.). *Intercultural Faultlines: Research Models in Translation Studies I. Textual and Cognitive Aspects*. Manchester: St. Jerome, 119–141.
- Meunier, Fanny, Sylviane Granger (Eds.) 2008. *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/z.138>
- Mukherjee, Joybrato, Sandra Götz (tulossa 2015). Learner corpora and learning context. – Sylviane Granger, Gaëtanelle Gilquin, Fanny Meunier (Eds.). *Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.
- Nesselhauf, Nadja 2005. *Collocations in a Learner Corpus*. Amsterdam: Benjamins. <http://dx.doi.org/10.1075/scl.14>
- Orpana, Terttu 1988. Kuvaus vai kommentti. Tutkimus suomen kielen adjektiiviadverbien semanttisesta tulkinnasta [‘Description or Comment? A Study on the Semantic Interpretation of Adjectival Adverbs in Finnish’]. *Opera Fennistica & Linguistica* 2. Tampere: Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
- Paradis, Carita 1997. *Degree Modifiers of Adjectives in Spoken British English*. Lund: Lund University Press.
- Pérez-Paredes, Pascual, María Belén Díez-Bedmar 2012. The use of intensifying adverbs in learner writing. – Yukio Tono, Yuji Kawaguchi, Makoto Minegishi (Eds.). *Developmental and Crosslinguistic Perspectives in Learner Corpus Research*. Amsterdam: John Benjamins, 105–124.
- Pérez-Paredes, Pascual, María Sánchez-Tornel 2014. Adverb use and language proficiency in young learners’ writing. – *International Journal of Corpus Linguistics* 19 (2), 178–200. <http://dx.doi.org/10.1075/ijcl.19.2.02per>
- Ringbom, Håkan 2007. *Cross-linguistic Similarity in Foreign Language Learning*. Clevedon: Multilingual Matters.
- Sabaté i Dalmau, Maria, Hortènsia Curell i Gotor 2007. From ‘Sorry very much’ to ‘I’m ever so sorry’. *Acquisitional patterns in L2 apologies by Catalan learners of English*. – *Intercultural Pragmatics* 4 (2), 287–315.
- Schauer, Gila 2009. *Interlanguage Pragmatic Development: The Study Abroad Context*. London: Continuum.
- Schinke-Llano, Linda 1990. Can foreign language learning be like second language acquisition? The curious case of immersion. – Bill Van Patten, James F. Lee (Eds.). *Second Language Acquisition, Foreign Language Learning*. Clevedon: Multilingual Matters, 216–225.

- Scott, Mike 2008. Developing WordSmith. – *International Journal of English Studies* 8 (1), 95–106.
- Scott, Mike, Christopher Tribble 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. SCL 22. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/scl.22>
- Segalowitz, Norman, Barbara F. Freed 2004. Context, contact, and cognition in oral fluency acquisition. Learning Spanish in at home and study abroad contexts. – *Studies in Second Language Acquisition* 26 (2), 173–199. <http://dx.doi.org/10.1017/S0272263104262027>
- SKTK = Suomen kielen tekstikokoelma. CSC – Tieteen tietotekniikan keskus. <https://sui.csc.fi>.
- Stenström, Anna-Brita 1999. He was really gormless, she's bloody crap: girls, boys and intensifiers. – Hilde Hasselgård, Signe Oksefjell (Eds.). *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam: Rodopi, 69–78.
- Takahashi, Tomoko, Leslie M. Beebe 1987. The development of pragmatic competence by Japanese learners of English. – *JALT Journal* 8, 131–155.
- VISK = Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen, Irja Alho 2004. *Iso suomen kielioppi* ['The large grammar of Finnish']. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkko-versio, viitattu 13.1.2015. <http://scripta.kotus.fi/visk> (16.12.2014).
- Virtanen, Veera 2011. *Minä lienen tullut joskus Suomessa vielä? Venäjänkielisten suomi toisena ja vieraana kielenä -oppijoiden perfektin ja imperfektin omaksumisen ongelmista*. Julkaisematon pro gradu -tutkielma. Jyväskylän yliopisto.
- Yaguchi, Michiko, Yoko Iyeiri, Yasumasa Baba 2010. Speech style and gender distinctions in the use of very and real/really: An analysis of the Corpus of Spoken Professional American English. – *Journal of Pragmatics* 42 (3), 585–597. <http://dx.doi.org/10.1016/j.pragma.2009.08.002>

Learning context and its effect on learner pragmatics: Degree modifiers as lexical teddy bears

JARMO HARRI JANTUNEN

University of Jyväskylä

The article focuses on the role of learning context in language acquisition. Learning context here refers to environments in which languages are learned: either a foreign language environment (often isolated from the target society, culture and language; mostly in a classroom and educational setting and consisting of formal learning) or a second language environment (surrounded by the target language, culture and society; often a natural setting outside of the classroom and comprising informal learning). Previous studies have shown that educational settings have an effect on how language is learned and which skills (e.g. pragmatic or grammatical competence) are mastered earlier.

The data for this study come from four corpora. There were two learner corpora: the International Corpus of Learner Finnish (ICLFI), consisting of foreign language data (texts produced by learners studying Finnish outside Finland); and the National Certificate Corpus (NCC), consisting of second language data (texts produced for proficiency tests in Finland). Both sets of data are rated according to the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001). The data for the present study comprise texts rated at level B1. The size of the ICLFI B1 data is 102,000 tokens and the NCC B1 data 23,500. In addition, two native language corpora and Internet data are used in the study.

Keyword analysis showed that certain degree modifiers are overused in the NCC compared to ICLFI, which indicates that the learning context may affect learner production. The results, however, seem to be contradictory as well as more complex than they appear. They support the conclusions of previous studies that suggest learners overuse degree modifiers in general. However, the range of degree modifiers seem to be more restricted in the second language production data than in the foreign language data. Furthermore, both learner groups tend to favour certain lexical teddy bears, but these differ according

to the learner group and learning context. Finally, the study reveals that the usage of degree modifiers is related to syntagmatic associations (collocations) and situational context (spoken vs. written language) and that the associations deviate from lexical structures produced by native speakers.

Keywords: learner language; learning context; corpora; keyword analysis; n-grams; degree modifiers

Jarmo Harri Jantunen

Jyväskylän yliopiston kielten laitos

PL35

FI-40014 Jyväskylä, Finland

jarmo.h.jantunen@ju.fi