



Turun yliopisto
University of Turku

EDISTYNEEN OPPIJANSUOMEN KONSTRUKTIOPIIRTEITÄ KORPUSVETOISESTI: AVAINRAKENNEANALYYSI

Ilmari Ivaska

Turun yliopisto

Humanistinen tiedekunta

Kieli- ja käännöstieteiden laitos

Suomen kieli ja suomalais-ugrilainen kielentutkimus

Tohtoriohjelma Utuling

Työn ohjaajat:

Lehtori, dosentti Kirsti Siitonen
Suomen kieli ja Suomalais-Ugrilainen
kielentutkimus
Turun yliopisto
Suomi

Professori Jarmo Jantunen
Suomen kieli
Jyväskylän yliopisto
Suomi

Tarkastajat:

Professori Maisa Martin
Suomen kieli
Jyväskylän yliopisto
Suomi

Professori Jaakko Leino
Suomen kieli
Helsingin yliopisto
Suomi

Vastaväittäjä:

Professori Maisa Martin
Suomen kieli
Jyväskylän yliopisto
Suomi

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck-järjestelmällä.

ISBN 978-951-29-6203-7 (PRINT)

ISBN 978-951-29-6204-4 (PDF)

ISSN 0082-6995

Painosalama Oy – Turku, 2015

TURUN YLIOPISTO

Humanistinen tiedekunta

Kieli- ja käännöstieteiden laitos

Suomen kieli ja suomalais-ugrilainen kielentutkimus

IVASKA, ILMARI: Edistyneen oppijansuomen konstruktiopiirteitä korpusvetoisesti:

avainrakenneanalyysi

Väitöskirja, 94 s., 134 liitesivua

Tohtoriohjelma Utuling

Syyskuu 2015

Ensikielen jälkeen opittavan kielen tutkimusta ja suomi toisena kielenä -alaa sen osana ovat koko niiden olemassaolon ajan hallinneet samat peruskysymykset: millaista oppiminen on eri vaiheissa ja eri ympäristöissä, sekä mikä oppimisessa on yleistä ja toisaalta mikä riippuu opittavasta kielestä ja oppijoiden kielitaustasta. Sähköisten oppijankielen tutkimusaineistojen eli korpusten lisääntymisen myötä tutkijat voivat aiempaa helpommin tutkia näitä ilmiöitä määrällisesti ja tarkastella oppijankielen sisäistä vaihtelua ja sen suhdetta tyypilliseen ensikieliseen kielenkäyttöön kielen eri osa-alueilla käyttöpohjaisesti eli todelliseen kielenkäyttöön pohjautuen. Tekninen kehitys on tuonut mukanaan aineisto- eli korpusvetoisuuden kaltaisia uusia tapoja lähestyä tutkimusaineistoa, jolloin tyypillisiä tutkimuskysymyksiä ”Miksi?” ja ”Miten?” edeltää kysymys: ”Mikä?”.

Tässä väitöskirjassa tarkastellaan edistyneiden suomenoppijoiden kirjoitettua akateemista kieltä ja suhteutetaan suomen oppimiselle ominaisia seikkoja käyttöpohjaisen mallin perusolettamuksiin. Aineisto on suomea toisena kielenä käyttävien opiskelijoiden tenttivastauksia, ja se on osa Edistyneiden suomenoppijoiden korpusta. Tutkimus on osin metodologinen, sillä väitöskirjassa esitellään ja siinä sovelletaan uutta korpusvetoista avainrakenneanalyysi-menetelmää, jonka avulla aineistoa lähestytään ilman hypoteeseja siitä, mitkä kielen ilmiöt ovat ominaisia edistyneelle oppijansuomelle. Tutkimus kuuluu kieliopin tutkimuksen piiriin, ja se nojaa kognitiivisen konstruktiokieliopin ajatukseen abstraktiudeltaan vaihtelevista konstruktioista kielijärjestelmän perusyksiköinä.

Tulokset puoltavat menetelmän sovellettavuutta kielen oppimisen tutkimukseen, sillä sen avulla kyettiin tunnistamaan konstruktioita, jotka erottavat edistyneitä oppijoita ensikielisistä kirjoittajista (esim. modaaliset verbiketjut), eri ensikieliä puhuvia suomenoppijoita (esim. konjunktiot) sekä konstruktioita, joiden käyttö muuttuu ajan kuluessa (esim. preteriti ja preesens). Monet havaitut erot ovat akateemisen kirjoittamisen erityispiirteitä, mikä tukee ajatusta kielen käyttö- ja kontekstikohtaisesta oppimisesta. Tuloksia voidaan yhtäältä soveltaa akateemisen kielitaidon opetuksessa. Toisaalta menetelmää voidaan käyttää kielenoppimisen tutkimuksen ohella uusien näkökulmien kartoittamiseksi erilaisten tai eri-ikäisten tekstien tyypillisten ominaisuuksien ja erojen tutkimuksessa.

Asiasanat: suomi toisena kielenä, akateeminen kirjoittaminen, korpuskielitiede, konstruktiokielioppi, käyttöpohjainen malli, avainrakenneanalyysi

UNIVERSITY OF TURKU

Faculty of Humanities

School of Languages and Translation Studies

Finnish and Finno-Ugric Languages

IVASKA, ILMARI: Corpus-driven Approach towards Constructional Features of Advanced Learner Finnish: Key Structure Analysis

Doctoral dissertation, 94 pp, 134 pp appendices

Doctoral Programme Utuling

September 2015

Research on second language acquisition and on Finnish as a second language has always been dominated by the same core questions: what is learning like in different phases and environments, which parts of learning are general and which parts are specific to the language being learned and to the learners' language background. Increasing availability of electronic learner language corpora facilitates quantitative and usage-based studies of these phenomena and of the relationship between first and second language use in different domains. Technical development has made it possible to approach the data from novel perspectives, so that the traditional research questions "Why?" and "How?" are preceded by a more general question "What?".

In this dissertation, I study academic Finnish written by advanced learners of Finnish and link phenomena specific to learning Finnish with the general assumptions of the usage-based model. The data are exam essays written by users of Finnish as a second language, and they are part of the Corpus of Advanced Learner Finnish. The study is in part methodological, as I introduce and apply a new corpus-driven method called Key Structure Analysis, which is designed to make it possible to approach data without prior hypotheses on which linguistic features are specific to advanced learner Finnish. The study is concerned with grammar, and it relies on the concept of constructions as the core units of linguistic system, as defined in cognitive Construction Grammar.

Results support the applicability of the method in language learning studies, as I was able to detect and define constructions that distinguish advanced learners from native writers (e.g. constructions of modality), constructions that distinguish learners of Finnish with different language backgrounds (e.g. conjunctions), and constructions whose use changes during longitudinal observation (e.g. the preterite and the present tense). Many of the observed differences are special characteristics of academic writing, which supports the idea that language learning is a usage- and context-based process. On the one hand, the results can be used in academic language teaching. On the other hand, the method introduced can be applied not only in language learning but also in bringing new perspectives to studies focusing on similarities and disparities between different kinds of texts or texts of different age.

Key words: Finnish as a second language, academic writing, corpus linguistics, Construction Grammar, usage-based model, key structure analysis

ALKUSANAT

Kielitieteellisen väitöskirjan kirjoittaminen voi olla yksinäistä puuhaa. Varmaa kuitenkin on, että yksin tämänkään väitöskirjan tekemisestä ei olisi tullut yhtään mitään. Ensimmäisenä haluan ehdottomasti kiittää ohjaajiani Kirsti Siitosta ja Jarmo Jantusta. Kirsti on viitoittanut tutkijanpolkuani sen ensi metreistä – kandiseminaarista alkaen – enkä kykene erittelemään kaikkia kiitollisuutta vaativia asioita. Kirstin roolit ovat vaihdelleet ohjaajasta esimieheen ja opettajakollegasta kanssakirjoittajaan ja paikat yhteistyölle toimistoista laivaterminaaleihin ja oikeastaan mihin tahansa, missä aurinko paistaa. Roolista ja paikasta riippumatta parhaita oppejani Kirstiltä on se, että maailmaa saa ja pitää muuttaa, mutta edellytyksenä on oltava perusasioiden hallinta. Kirstillä on myös uskomaton taito oikeiden kysymysten esittämiseen umpikujista ulospääsemiseksi. Jarmon tapasin ensi kerran pitäessäni elämäni ensimmäistä akateemista esitelmää *Korpusuuringute metodoloogia ja märgendamise probleemid* -seminaarissa Tallinnassa syksyllä 2008, minkä jälkeen hän on sekä asiantuntevan ohjauksensa että omien julkaisujensa kautta auttanut minua ymmärtämään korpuskielitieteen perusteita, traditiota ja mahdollisuuksia. Jarmon rakentava ja tarkkanäköinen palaute on kerta kerran jälkeen osoittanut tutkimuksessani huomiota vaativia seikkoja, joille olin jo itse sokeutunut.

Esitarkastajiltani en olisi voinut toivoa enempää: professori Maisa Martin on perustanut tieteenalan, jolle tutkimukseni sijoittuu, ja professori Jaakko Leino on soveltamani teoreettisen viitekehyyksen ehdoton eksperti. Artikkelimuotoisessa väitöskirjassa tilaa muutoksille on luonnollisesti vähän, mutta saamani lausunnot kirkastivat useita seikkoja ja auttoivat minua hahmottamaan tutkimukseni aseman entistä paremmin.

Turun yliopiston suomen kielen ja suomalais-ugrilaisen kielentutkimuksen oppiaine on parhaita mahdollisia akateemisia kotipesiä. Asenne uusia ajatuksia kohtaan on avoin, ja monenlaista virallista ja todellista tukea on ollut saatavilla niin ylhäältä päin kuin rinnaltakin, mistä kiitän syvästi kaikkia asianosaisia. Haluan lisäksi erikseen kiittää Kaisa Häkkistä kaikesta tuesta ja ymmärryksestä ajoittaiseen etätyöskentelyyn. Marja-Liisa Helasvuo on neuvonut ja inspiroinut minua haastamaan itseni niin fennistinä kuin lingvistinäkin ja kannustanut uusien ajatusten pariin niin lähelle kuin kauas, ja Tuomas Huumo on osoittanut, miten kiehtova suomi voi olla myös yleistypologisesta näkökulmasta – kiitos teille molemmille. Nobufumi Inaba on väsymättä näyttänyt, korjannut, opettanut ja selvittänyt minulle, miten korpuksia tehdään ja mitä tarkoittaa systemaattisuus, ja Tommi Kurki on toistuvasti jaksanut keskustella kanssani aineistoista, metodeista, laskemisesta ja alasarjajalkapallosta – suurkiitos molemmille, ja varaan oikeuden avata Nobulle oven ainakin kerran lähitulevaisuudessa.

Laura Tyysterin kanssa olemme korkeakouluharjoittelija-ajoista lähtien opetelleet yhdessä tieteen tavoille vaihtelevalla menestyksellä – polun tallaamisesta yhdessä on ollut suunnaton apu, vaikka reittimme ovatkin aika ajoin poikenneet toisistaan. Lisäksi EILC-kurssit useana kesänä ovat pitäneet kontaktin kielenoppijoiden ensiaskeleisiin kirkkaana mielessä ja

erakoituvan tutkijan järjen päässä, mistä kiitos Elina Aholalle, Hanna Jokelalle, Leena Maria Heikkolalle ja Rosa Salmelalle. Hannalle lisäksi erityiskiitos: olet paras tuntemani sohvamiesnainen! Teitä kanssakulkijoita on liikaa erikseen nimeltä kiitettäväksi vaikka kaikki sen ansaitsisivatkin, ja kaikkien kanssa olen käynyt hedelmällisiä keskusteluja niin kielestä kuin ruiskakostakin – kiitos ja kumarrus!

Aloitin väitöskirjan teon SKR:n Varsinais-Suomen rahaston tuella, minkä jälkeen pääsin osaksi valtakunnallista kieliaineiden tohtoriohjelmaksi Langnetia, ja kiitänkin molempia tahoja luottamuksesta ja tutkimukseni taloudellisesta mahdollistamisesta. Myös Turun yliopistosäätiö on myötävaikuttanut rahoittamalla Edistyneiden suomenoppijoiden korpus-hanketta, josta tämäkin väitöskirja ponnistaa. Langnetia ja sen Kielen oppiminen -alaohjelmaa minun on lisäksi kiittäminen kiinnostavista kursseista, lukemattomista hyvistä hetkistä ja Satu Rakkolainen-Sossaa lisäksi mahtavasta matka- ja manausseurasta.

Olen kotonani väliaikaisuudessa, ja myös tämän väitöskirjan ajatuksia ja osia on työstetty ainakin kuudessa maassa ja kahdella manterella, niin maalla, ilmassa kuin vedessäkin. Moni paikka palermolaisista vierasmajoista ballardilaisiin ullakkohuoneisiin on koitunut tälle väitöskirjalle tärkeäksi omalla tavallaan, mutta niiden läpikäyminen ei tässä ole mahdollista. Erikseen haluan kuitenkin kiittää nykyistä työnantajaani University of Washingtonia ja erityisesti professori Andrew Nestingeniä tutkimusmyönteisestä asenteesta ja tuesta. Yksi puolivuotinen kirjoitusretiriitti kehkeytyi vaimoni harjoittelujakson myötä Ateenaan, ja kiitänkin siksi Suomen Ateenan-instituuttia ja sen silloista johtajaa Martti Leiwoa ymmärryksestä, joustavuudesta ja yösijasta.

Ystäviäni haluan kiittää kaikesta siitä, mikä vie ajatukset pois tutkimuksesta; olipa se sitten musiikkia, ruokaa jalkapalloa tai mitä tahansa muuta. Perhettäni en olisi osannut paremmin valita: äiti, isä ja Lauri sekä Lotta, Unto ja Valto – kiitos kaikesta kannustuksesta ja tuesta! Kaikkein suurin kiitos kuuluu rakkaalle vaimolleni Lauralle: koska Lauran kanssa on normaalia poimia aamupalan aikana kirjahyllystä vielä viidennen eri kielen kielioppi ja tarkistaa jokin tuiki tärkeä verbintaivutusparadigma, koska Laura on muistuttanut, että satunnaismetsien kasvaessa oven alta uniin on aika tehdä jotain muuta ja koska Laura on tukenut jokaista askeltani.

Tien päällä, elokuun 3:ntenä päivänä vuonna 2015,

Ilmari Ivaska

Sisällysluettelo

Luettelot taulukoista ja kuvioista	2
1. Aluksi.....	3
2. Käyttöpohjainen näkemys kielen rakenteisiin ja niiden oppimiseen	7
2.1. Lauserakenteen kuvaaminen kielen oppimisen kannalta	7
2.1.1. Analogian ohjaama käyttöpohjainen malli	8
2.1.2. Konstruktiot kielen rakenteen kuvaajina ja opittavina ilmiöinä	11
2.1.3. Fraseologisten yksiköiden ja konstruktioiden suhde	16
2.1.4. Kohteena konstruktiot – lähtökohtana toistuvat mallit	20
2.2. Toisen kielen oppiminen.....	29
2.2.1. Ensimmäisen ja toisen kielen välinen suhde	31
2.2.2. Konstruktiot toisen kielen opittavina yksiköinä.....	34
2.2.3. Oppijakohtaiset tekijät	39
2.2.4. Tekstilaji käyttökontekstin kuvaajana	42
3. Avainrakenneanalyysi korpustutkimuksen kentällä	49
3.1. Korpus oppijankielen tutkimusaineistona.....	49
3.2. Korpusvetoisuus ja tilastollinen avaimuus	50
3.3. Deskriptio ja kontrastointi oppijankielen korpustutkimuksessa	52
4. Tämän tutkimuksen aineisto ja menetelmät	58
4.1. Edistyneiden suomenoppijoiden korpus LAS2	58
4.2. Avainrakenneanalyysi.....	61
5. Tulokset.....	64
5.1. Määrällinen yleiskuva havaituista potentiaalisista avainmuodoista.....	64
5.2. Edistynyttä oppijansuomea ensikielisestä suomesta erottavat konstruktiot	66
5.3. Ensikielikohtaiset erot edistyneen oppijansuomen konstruktioissa	70
5.4. Käyttöpohjainen näkökulma eroihin ja muutoksiin	73
5.4.1. Mahdollisuuden ilmaukset	73
5.4.2. Aikamuodot muutoksessa.....	75
6. Yhteenvedo ja päätelmät	78
Lähteet.....	83
Alkuperäiset julkaisut.....	95

Osatutkimukset:

1. The Corpus of Advanced Learner Finnish (LAS2): Database and toolkit to study academic learner Finnish. *Apples – Journal of Applied Language Studies* 8(3) s. 21–38 (2014). Tekstissä *LAS2*.
2. Edistyneen oppijansuomen avainrakenteita. Korpusnäkökulma kahden kielimuodon tyypillisiin rakenteellisiin eroihin. *Virittäjä* 118 s. 161–193 (2014). Tekstissä *S2-avainrakenteet*.
3. 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* s. 23–44 (2015). Tekstissä *Siirtovaikutus-avainrakenteet*.
4. Longitudinal changes in academic Learner Finnish: A key structure analysis. *International Journal of Learner Corpus Research* 1(2) s. 210–241 (2015). Tekstissä *Pitkittäismuutokset*.
5. Mahdollisuuden ilmaiseminen S1-suomea ja edistynyttä S2-suomea erottavana piirteenä. *Lähivertailuja. Lähivördlusi* 24, s. 47–80 (2014). Tekstissä *Mahdollisuuden ilmaukset*.

LUETTELOT TAULUKOISTA JA KUVIOISTA

Taulukko 1. Väitöskirjan osatutkimusten aineistot. S1-aineiston kaikki tekstit ovat eri kirjoittajilta.....	61
Taulukko 2. S1- ja S2-aineistoja toisistaan parhaiten erottavat n-grammit ja niiden frekvenssien keskiarvot. Grammit ovat taulukossa pituuden mukaan järjestettynä 1-grammeista alkaen, ja samanmittaiset grammit ovat siinä järjestyksessä, miten hyvin ne erottavat aineistoja toisistaan.....	65
Taulukko 3. Eri ensikielisten kirjoittajien tekstejä toisistaan eniten erottavat n-grammit ja niiden frekvenssien keskiarvot. Vertailuaineiston (fi) lisäksi tarkastellut ensikielet ovat tšekki (cs), unkari (hu), japani (ja), liettua (lt) ja venäjä (ru).....	66
Taulukko 4. Finiittimuotoisen indikatiiviverbin preteritin yksikön kolmannen persoonan (<fin ind pret sg3>) frekvenssin keskiarvot 16 kuukauden tarkastelujakson eri vaiheissa ja vertailuaineistossa.....	66
Kuvio 1. Noodin ja kollokaatin välinen kollokaatiosuhde.....	25
Kuvio 2. Yksittäisen kielen ilmiön oppimisprosessia kuvaava DEMfad-malli (Martin ym. 2010: 59).....	54

1. ALUKSI

Ensikielen jälkeen opittavan kielen eli ns. toisen kielen tutkimusta ja sitä kautta myös suomi toisena kielenä -tutkimusta (S2) ovat koko alojen olemassaolon ajan hallinneet monilta osin toistuvat peruskysymykset: mikä on oppijankielelle ominaista oppimisen eri vaiheissa, minkälaisia ovat kielenoppimisen yleiset mekanismit, miten opittava kieli vaikuttaa oppimisprosessiin ja miten oppijat ovat samanlaisia tai eroavat toisistaan esimerkiksi ensikielensä osalta. Kielellisten tekijöiden lisäksi myös kielenulkoisten tekijöiden sekä kielen oppimis- ja käyttökontekstin vaikutusta oppimiseen on pohdittu jo pitkään. (S2-tutkimuksen asemasta ja nykytilasta toisen kielen oppimisen tutkimuksessa ks. esim. Suni 2012.) Nykyisin kielijärjestelmä ja sen oppiminen mielletään usein niin toisen kielen tutkimuksessa kuin kielitieteessä ylipäätäänkin perusluonteeltaan käyttöpohjaiseksi niin, että käytetty kieli kuvastaa sellaisia todennäköisiä kielellisiä tapoja, jotka kehkeytyvät luonnollisen kielenkäytön ilmauksista. Kielenkäyttö ei siis ainoastaan kuvaa abstraktia kielijärjestelmää, vaan käytetty kieli muodostaa itse järjestelmän. Keskeisiä seikkoja kielijärjestelmälle ovat sen analoginen perusluonne, eli kielenkäytön toistuvat rakenteellis-funktionaalis-semanticiset samankaltaisuudet, kyky samankaltaisuuden tunnistamiseen sekä kyky rakentaa näiden samankaltaisuuksien varaan yleistyksiä. (Rosch & Mervis 1975; Hopper 1987; Bybee & Thompson 1997; ks. myös Skousen 1989; Itkonen 2005.)

Oppijankielen sähköisesti käytettävien korpusten lisääntyminen on mahdollistanut toisen kielen oppimista käsittelevien kysymysten lähestymisen osin uudella tavalla, kun tutkijoiden käyttöön on tullut suuria aineistoja, joiden avulla oppijankieltä on voinut entistä helpommin lähestyä myös määrällisestä näkökulmasta ja tarkastella tällä tavoin oppijankielen tyypillisyyksiä ja vaihtelua (ks. esim. Granger 2008). Korpusaineistojen avulla toisen kielen oppimista voidaan myös tarkastella suhteessa tyypilliseen ensikieliseen kielenkäyttöön ja siinä esiintyvään vaihteluun ylipäätään ja päästä sillä tavoin lähemmäs kielen käyttöpohjaista perusluonnetta. Näin oppijankieltä voidaan tarkastella ja kuvata sekä sen omista lähtökohdista että suhteessa ensikieliseen kielenkäyttöön neutraalin deskriptiivisesti ilman, että kummastakaan tehdään minkäänlaisia oletuksia niiden oikeellisuuden tai virheellisyyden osalta. (Granger 1996, 2015.)

Oppijankielen korpukset mahdollistavat paremmin myös edistyneen oppijankielen ja edistyneiden kielenoppijoiden tutkimuksen. Edistyneet kielenoppijat kykenevät toimimaan oppimallaan kielellään akateemisen ympäristön kaltaisissa, monipuolista ja abstraktia

kielitaitoa vaativissa tilanteissa. Tutkimuksessa on kuitenkin havaittu, että usein toisen kielen oppijat osaavat kieltä laadullisesti hyvin mutta että heidän kielelliset valintansa ovat käyttökontekstissaan epätyypillisiä (ks. esim. Bardowi-Harlig & Dörnyei 1998). Edistynyt kielitaito onkin myös tilannekohtaisten toimintatapojen tuntemusta – olipa kyseessä ensikieli tai ensikielen jälkeen opittu kieli (Swales 1990: 58; Hyland 2004: 54–56; Östman 2005). Laajan aineiston avulla kielen oppimista voi tarkastella laadullisen oikeakielisyyden ohella myös kielenpiirteiden käyttöfrekvenssin ja niiden funktioiden jakautumisen ja näissä tapahtuvien muutosten kannalta (Martin ym. 2010). Näin edistyneestä oppijankielestä saadaan tarkempi ja moniulotteisempi kuva, mikä puolestaan voi auttaa ymmärtämään sitä, mikä erottaa vielä edistyneitäkin kielenoppijoita ensikielisistä kielenkäyttäjistä ja mistä nämä erot johtuvat.

Kielikorpusten kasvava määrä sekä tietokoneiden ja tietokoneohjelmien kehityksen tuomat mahdollisuudet ovat vaikuttaneet myös niihin tapoihin, joilla kieltä voidaan tutkia. Yhtenä verraten uutena tapana jaotella erilaisia korpustutkimuksen lähestymistapoja on erottaa toisistaan korpus- tai aineistovetoinen tutkimusote. (Tognini-Bonelli 2001.) Silloin kun kielentutkimus perustuu todelliseen kieliaineistoon, tutkimusasetelman tyypillisiä piirteitä ovat tutkijan havaintoihin, kielelliseen intuitioon ja aiempaan tutkimukseen perustuva tutkimuksellisesti kiinnostavan ilmiön tunnistaminen ja määrittely, näihin perustuvan hypoteesin esittäminen ja tämän hypoteesin testaaminen (Eddington 2008). Myös korpusvetoinen tutkimusote seuraa tätä perusasetelmaa monin tavoin, mutta tutkittavaa kielenilmiötä ei kuitenkaan valita subjektiivisen havainnoinnin perusteella. Sen sijaan voidaan käyttää hyväksi suuren aineiston tarjoamia mahdollisuuksia ja laskennallisia menetelmiä sellaisten ilmiöiden tavoittamiseksi, joiden käytössä ilmenee jotakin kulloinkin käsillä olevan tutkimuksen kannalta mielenkiintoista. Annetaan siis aineiston johdattaa tutkija kiinnostavan tutkimuskysymyksen äärelle. Karrikoiden voitaneen sanoa, että tyypillisiä tutkimuskysymyksiä ”Miksi?” ja ”Miten?” edeltää tässä tapauksessa vielä yksi perustavanlaatuisempi kysymys: ”Mikä?”.

Tämä väitöskirjatutkimus pyrkii tarkastelemaan edistyneiden suomenoppijoiden kirjoitettua kieltä käyttöpohjaisesta näkökulmasta ja suhteuttamaan suomen kielen oppimiselle ja toisena kielenä käytetylle suomelle ominaisia seikkoja käyttöpohjaisen mallin perusolettamuksiin. Väitöskirjan tavoite on osin metodologinen: olen kiinnostunut siitä, miten oppijankieltä ja suomea toisena kielenä voidaan tutkia korpus- eli aineistovetoisesti siten, että tutkimuksen

alussa ei aseteta hypoteeseja siitä, mitkä kielen ilmiöt ovat ominaisia edistyneelle oppijansuomelle. Paneudun väitöskirjassani kielen rakenteellisiin ominaispiirteisiin, ja työni kuuluukin kieliopin tutkimuksen piiriin. Näkemykseni kielen rakenteiden luonteesta nojaa kognitiivisen konstruktioikieliopin ajatukseen todelliseen kielenkäyttöön perustuvista, skemaattisuudeltaan, monimutkaisuudeltaan ja laajuudeltaan vaihtelevista konstruktioista kielijärjestelmän perusyksiköinä (ks. esim. Goldberg 1995, 2006). Tarkastelunalaiset suomelle toisena kielenä tyypilliset konstruktioit voivat siis olla niin yksittäisiä morfologisia muotoja, sanoja, leksikaalisesti rajattuja monisanaisia rakenteita kuin lausetyypinomaisia täysin skemaattisia rakenteitakin – tai tällaisten tekijöiden yhteensulaumia. Väitöskirjan tutkimuskysymykset voidaan tiivistää seuraavasti:

- 1) Miten oppijankielen rakenteellisia ominaispiirteitä voidaan tunnistaa ja analysoida korpus- eli aineistovetoisesti – ilman ennalta asetettua hypoteesia mahdollisten erojen luonteesta?
- 2) Mitkä kielen konstruktioit ovat tyypillisiä tai epätyypillisiä edistyneiden suomenoppijoiden kirjoituksille ensikielenään suomea käyttävien kirjoituksiin verrattaessa?
- 3) Mitkä kielen konstruktioit erottavat eri ensikieliä puhuvia edistyneitä suomenoppijoita toisistaan?
- 4) Minkä kielen konstruktioiden käyttö muuttuu edistyneiden suomenoppijoiden kirjoituksissa ajan kuluessa ja miltä muutos näyttää ensikielenään suomea käyttävien kirjoituksiin verrattuna?
- 5) Mistä syistä vielä edistyneidenkin suomenoppijoiden kieli poikkeaa rakenteellisesti joiltakin osin ensikielenään suomea käyttävien kielestä ja on edelleen muutoksessa?

Pyrin vastaamaan näihin tutkimuskysymyksiin viiden artikkelina julkaistun osatutkimuksen avulla. Lähestyn tutkimuskysymyksiä kolmesta toisen kielen tutkimukselle tyypillisestä näkökulmasta: 1) miten toisen kielen käyttäjät eroavat ensikielisistä kielenkäyttäjistä, 2) miten eri ensikieliä puhuvat toisen kielen oppijat eroavat toisistaan ja 3) miten toisen kielen oppijoiden kieli muuttuu ajan kuluessa? Soveltamassani korpusvetoisessa tutkimusasetelmassa käytetyn aineiston luonteella on keskeinen osa, ja yksi osatutkimuksista paneutuukin yksinomaan käytetyn korpuksen ja sen mahdollistamien menetelmien tarkasteluun. Aineiston luonne ja sen käyttöön kehitetyt työkalut määrittävät ja rajoittavat sitä, minkälaisiin kysymyksiin sen avulla ylipäätään on mielekästä ja mahdollista vastata. Aineisto myös määrittää sitä, miten rakenteellisia ominaispiirteitä tunnistetaan korpusvetoisesti tämän tutkimuksen puitteissa.

Käytetyn menetelmän kaksivaiheisen luonteen vuoksi tarkastelen tapauskohtaisesti kullekin tutkimusnäkökulmalle relevanttien kielenpiirteiden paikantamisen myötä useiden havaittujen konstruktioaalisten erojen luonnetta ja mahdollisia syitä. Tarkempaan

tarkasteluun valikoituvat piirteet siis nousevat esiin kunkin osatutkimuksen kysymyksenasettelun ja siihen perustuvan määrällisen analyysin kautta. Esimerkiksi toisen kielen käyttäjiä ja ensikielisiä kielenkäyttäjiä erottavat seikat ulottuvat *esimerkiksi*-sanankaltaisten, leksikaalisesti spesifien konstruktoiden käyttöeroista aina hyvin abstraktien, erilaisten mahdollisuutta ilmaisevien konstruktoiden käyttöeroihin. Japania, unkaria tai viroa ensikielenään puhuvat S2-kielenkäyttäjät puolestaan käyttävät mm. konjunktioita huomattavasti enemmän kuin liettuaa, tšekkiä tai venäjää ensikielenään puhuvat S2-kielenkäyttäjät. Vastaavasti S2-kielenkäyttäjien aikamuotojen käyttö muuttuu pitkittäistarkastelussa siten, että preteritin määrä vähenee preesensin määrän kasvaessa. Pohdin lisäksi mahdollisia syitä havaittuihin eroihin ja pyrin tätä kautta linkittämään tutkimuksen yleisemmällä tasolla teoreettiseen keskusteluun kielen perusolemuksesta. Vastaamalla näihin kysymyksiin kuvaan samalla osana tätä väitöskirjaa kehittämäni tutkimusmenetelmän – avainrakenneanalyysin – toimintamekanismeja ja käytettävyyttä sekä periaatteen tasolla että soveltamalla sitä käytännön tutkimustyöhön, mistä syystä pohdin myös menetelmän käyttökelpoisuutta tätä tutkimusta yleisemmällä tasolla.

Tämän väitöskirjan johdanto-osion tavoitteet ovat kahtalaiset: ensinnäkin valotan sitä teoreettista viitekehystä, johon tämä tutkimus nivoutuu. Kuvaan luvussa 2 sitä, miten toisen kielen oppiminen ja nyt käsillä oleva tutkimus asemoituvat osaksi käyttöpohjaisten mallien kirjoa. Luvussa 3 käsittelen korpusten roolia ja korpusten tarjoamien menetelmien vaikutusta toisen kielen oppimisen tutkimukseen ja luvussa 4 kuvaan tämän tutkimuksen aineiston ja käytetyt menetelmät. Luvussa 5 luon yleiskuvan osatutkimuksissa saatuihin tuloksiin ja niiden mahdollistamiin tulkintoihin sekä suhteutan tuloksia aiempaan tutkimukseen. Luvussa 6 kokoaan lopuksi yhteen tekemäni päätelmät suhteessa esittämiini tutkimuskysymyksiin ja pohdin lyhyesti tämän tutkimuksen vaikutuksia tulevaan tutkimukseen.

2. KÄYTTÖPOHJAINEN NÄKEMYS KIELEN RAKENTEISIIN JA NIIDEN OPPIMISEEN

2.1. Lauserakenteen kuvaaminen kielen oppimisen kannalta

Tämän tutkimuksen keskiössä ovat kielen rakenteelliset piirteet ja rakenteessa ilmenevät erot ensimmäisen kielen jälkeen opitun suomen ja ensikielisen suomen välillä. Tutkimus nojaa käyttöpohjaiseen kielikäsitteeseen, enkä erota kielikykyä ja kielen tuottamista toisistaan perustavanlaatuisesti erillään oleviksi osa-alueiksi. En myöskään tee kategorista eroa kielen oppimisen (engl. *learning*) ja omaksumisen (engl. *acquisition*) välillä, vaan käytän termiä oppiminen kuvaamaan kielitaitoa ja sen muutoksia riippumatta esimerkiksi sen neurologisen prosessin luonteesta, jolla kieltä käytetään (ks. yhteenvetona esim. Ellis, N. 2002). Kuvaan tässä alaluvussa sitä teoreettista viitekehystä, jonka puitteissa tarkastelen kielen rakennetta – niin oppimista kuin käyttöäkin. Toisella kielellä toimimiseen ja sen oppimiseen liittyviä erityispiirteitä tarkastelen luvussa 2.2.

Kaikki nyt käsiteltävät näkemykset jakavat ajatuksen kielen käyttöpohjaisesta luonteesta eli siitä, että kielessä ilmenevät rakenteelliset samankaltaisuudet kumpuavat luonnollisesta kielenkäytöstä. Kuvatesani kielen rakenteiden luonnetta käsitteelen sekä konstruktioita että fraseologisia merkitysyksiköitä. Terminologisesta erosta huolimatta molemmilla käsitteillä voidaan nähdäkseni kuvata samaa kielijärjestelmän ilmiötä – sitä, miten yksittäisiä sanoja laajemmat kokonaisuudet tyypillisesti käyttäytyvät, miten muoto ja merkitys tyypillisesti yhdistyvät todellisessa kielenkäytössä ja miten tällaisia muoto–merkitys-kokonaisuuksia opitaan. Näkemysten painotus on kuitenkin hieman erilainen. Nähdäkseni painotukset kuvaavat lähinnä sitä kielen osa-aluetta, johon kuvaustavat ovat ensi sijassa keskittyneet; konstruktioista puhuttaessa tutkimuksen fokus on usein lausetason syntaktisissa ilmiöissä, kun taas fraseologinen tarkastelu on pitkään niveltynyt ennen muuta leksikografiaan ja leksikologiaan. Monissa tapauksissa lähestymisten välinen rajankäynti on kuitenkin teennäistä eikä niiden perustavanlaatuisen erottaminen ole nähdäkseni mielekästä: molemmissa näkemyksissä odotetaan vähintään implisiittisesti tarkasteltavien piirteiden muodostavan semanttisesti koherentin merkitysyksikön. Lisäksi nämä muoto–merkitys-parit määrittyvät kielenilmausten rakenteellisen samankaltaisuuden ja sen kanssa korreloivan samankaltaisen merkityksen avulla, eli sen perusteella, että tietynmuotoinen kielenilmaus esiintyy toistuvasti ilmaisemassa samaa tai samankaltaista merkitystä. Kuvaukset perustuvat siis kielen analogiseen luonteeseen eli siihen, että kieli sekä järjestyy että toimii rakenteellisten samankaltaisuuksien avulla. Toistuvat mallit sen sijaan kuvaavat samanlaisina

toistuvia kielen sekvenssejä laskennalliselta kannalta riippumatta siitä, muodostaako toistuva malli minkäänlaista koherenttia merkityskokonaisuutta. Tarkastelemalla havaittavien toistuvuuksien ominaisuuksia huomio voidaan kuitenkin keskittää konstruktioomaisiin kokonaisuuksiin tai fraseologisiin merkitysyksiköihin, mistä syystä kuvaan tässä luvussa niitä käsitteitä, joiden avulla tarkastelen tätä muodon ja merkityksen välistä yhteyttä.

2.1.1. Analogian ohjaama käyttöpohjainen malli

Kielen käyttöpohjaiset mallit eivät muodosta mitään yhtenevää ja koherenttia teoreettista näkemystä kielestä. Niitä kuitenkin yhdistää laajasti katsoen se, että kielikyvyn (engl. *competence*) ja tuottamisen (engl. *performance*) välille ei tehdä mitään perustavanlaatuista eroa (vrt. Chomsky 1965: 4) ja että kieliopin teorian kannalta ne jakavat tyypillisesti kognitiivisen kielitieteen näkemyksen kielen luonteesta (Langacker 1987; sovellus suomeen, ks. P. Leino 1993). Nämä oletukset ovat myös tämän väitöskirjatutkimuksen teoreettisena ytimenä, ja näkemykseni kielen luonteesta nivoutuukin osaksi käyttöpohjaisten mallien kirjoa. Käyttöpohjainen kielinäkemys ei suinkaan ole 1900-luvun viimeisten vuosikymmenten keksintöä, sillä esimerkiksi Hermann Paul kehitti jo 1800- ja 1900-luvun vaihteessa hyvin kattavaa käyttöpohjaista kieliteoriaa (Paul 1891). Paul toteaa kielentutkimuksen todellisen kohteen – luonnollisen kielen – olevan kielenkäyttäjien tuottamien ilmausten koko joukon (mts. 2–3). Paul kuvaa edelleen kielenilmausten järjestyvän muoto–merkitys-pareiksi ja samankaltaisten parien välisiksi assosiaatioiksi niiden todellisen käytön perusteella ilman tietoista toimintaa, eikä niitä tule täten nähdä itsessään kieliopillisina kategorioina, jotka puolestaan ovat tietoisien abstrahoinnin tulosta kielen ilmiöiden kuvaamiseksi (mts. 5–6). Kuten Elisabeth Traugott ja Graeme Trousdale huomauttavat, generatiivinen kielioppitraditio kuitenkin jätti käyttöpohjaisen ajattelun monilta osin varjoonsa 1900-luvun puolivälistä aina 1980-luvulle asti (Traugott & Trousdale 2013: 46–47).

Käyttöpohjaisia malleja yhdistää se, että kielen ajatellaan koostuvan todellisten kielellisten ilmausten kokonaisuudesta sekä näiden ilmausten määrällisistä ja rakenteellis-funktionaalisista suhteista. Kielioppi ja kielen psykologinen todellisuus ovat siis tässä valossa joukko todennäköisiä kielellisiä tapoja, jotka kehkeytyvät (engl. *emerge*) todellisen kielenkäytön ilmauksista ilmausten samankaltaisuuden ja erilaisuuden ja niiden varaan rakentuvien yleistysten perusteella (Hopper 1987; Bybee & Thompson 1997). (Yleiskuva käyttöpohjaisista malleista ja niiden keskinäisistä suhteista ks. esim. Barlow & Kemmer 2000; Bybee 2010.) Seuraten tätä ajattelutapaa esimerkiksi Diane Larsen-Freeman kuvaa kielen

strukturoivan itseään alhaalta ylöspäin siten, että jokainen käytetty kielen ilmaus muuttaa kielen kokonaisuutta. Kielijärjestelmä kehkeytyy synteettisesti sen alajärjestelmien (fonologia, morfologia, leksikko, syntaksi, semantiikka, pragmatiikka) välisessä kanssakäymisessä, mistä syystä kunkin alajärjestelmän yksityiskohtainen kuvaaminen ei vielä kata kielen koko kuvaa. (Larsen-Freeman 1997: 147–149.) Joan Bybee ja Paul Hopper puolestaan toteavat kielen käyttöyksiköiden (engl. *units of usage*) olevan suurelta osin myös todellisia varastointiyksiköitä (engl. *units of storage and access*). Näiden yksiköiden oletetaan puolestaan olevan usein monisanaisia konstruktioomaisia joukkoja, joissa on vaihteleva määrä vapaita, rajatusti vapaita tai rajattuja paikkoja. (Bybee & Hopper 2001: 8–9.) Kielijärjestelmän yksiköt ja niiden muodostamat kategoriat ovat siis luonteeltaan käyttöpohjaisesti abstrahoituneita joukkoja, jotka määrittyvät suhteessa muihin yksiköihin ja kategorioihin niiden funktionaalisen ja rakenteellisen samankaltaisuuden perusteella siten, että kullakin kategorian jäsenellä ei tarvitse välttämättä olla mitään yksittäistä yhteistä nimittäjää (Rosch & Mervis 1975; ks. myös 2.1.2).

On selvää, että kielenkäyttäjät tuottavat jatkuvasti myös sellaisia ilmauksia, joita he eivät ole koskaan kuulleet. Tätä ilmiötä kutsutaan yleensä joko analogiaksi tai yleistykseksi. Esa Itkonen toteaa analogian olevan kahden kokonaisuuden välillä olevaa rakenteellista tai funktionaalista samankaltaisuutta, joka esiintyy kielen eri abstraktiotasoilla ja myös kielenulkoisissa kokonaisuuksissa (Itkonen, Esa 2005). Analogisen tarkastelun avulla voidaankin luontevasti jäsentää sitä, mikä on joillekin kielenpiirteille¹ yhteistä tai mikä on jonkin kielenpiirteen suhde johonkin toiseen kielenpiirteeseen. Adele Goldberg kuvaakin juuri analogian keinoin kielenpiirteiden välisiä suhteita periytyvyshierarkiaksi kutsumansa mekanismin avulla, jonka mukaan tietyn rakenteen ominaisuudet siirtyvät sen alapuolella oleville rakenteille (Goldberg 2006: 13–14), mikäli alapuolella olevat rakenteet eivät erikseen poikkea tästä yläkategoriasta ja näin muodosta erillistä alakategoriaansa (Goldberg 1995: 108). Kielen analogiaan perustuva funktionaalisuus ja produktiivisuus ovat luontevia näkökulmia nimenomaan käyttöpohjaisten kieliteorioiden viitekehyksessä. Esimerkiksi Royal Skousen kuvaa analogisessa mallissaan kielen rakenteellisten oletusten perustuvan nimenomaan aiemmin kohdattuihin todellisiin ilmauksiin ja niistä muodostuvien samankaltaisten esiintymien joukkoihin ja näiden joukkojen distribuutioihin (Skousen 1989: 22–23). Nämä samalla tavalla käyttäytyvät esiintymät puolestaan voidaan nähdä

¹ Käytän tässä väitöskirjatutkimuksessa termiä kielenpiirre hyvin geneerisenä yleiskäsitteenä mille tahansa kielelliselle yksikölle foneemeista morfeemeihin, lekseemeihin ja aina lauseisiin ja virkkeisiin saakka riippumatta niiden merkityksestä, funktiosta tai muodon ja merkityksen välisestä koherenssista.

kielenpiirteiden joukkoina, joita kieliopin kirjoituksessa yleensä käsitellään kategorioina tai paradigmoina. Tämä kuva analogiasta vastaa hyvin käyttöpohjaisissa malleissa kuvattavaa kielen rakenteiden skemaattisuutta ja sen asteittaista kehittymistä todellisessa kielenkäytössä havaittujen frekvenssien pohjalta (ks. esim. Croft & Cruse 2004: 308–312).

Rakenteellisen samankaltaisuuden hahmottamista ja soveltamista voidaan pitää jatkuvan kielenkäytön lisäksi myös kielenoppimisen keskeisenä mekanismina (ks. esim. Tomasello 2003: 4–5, 144; Laalo 2011: 14–18). Samankaltaisuuden ja erilaisuuden havaitseminen mahdollistaa niin muodoltaan kuin kontekstuaalisestikin samankaltaisten kielen jaksojen (sanojen, lausekkeiden, lauseiden jne.) kategorioimisen, ja käyttöpohjaisiin malleihin sisältyykin tyypillisesti vähintään implisiittisesti myös oletus kielen käyttöpohjaisesta oppimisesta (yleiskuvauksena käyttöpohjaisista malleista kielenoppimisen selittäjinä, ks. Tomasello 2003). Avainseikkana on tällöin se, että rakennetta ei opita abstrakteina sääntöinä tai universaalien parametrien sääntöinä. Kielen kompositionaaliset ominaisuudet koostuvat oppijan ympäristöstään saamasta syötöksestä ja syötöksen yksittäisten ilmausten kautta leviävästä systemaattisuudesta (Hopper 1987: 142). Näkemystä kutsutaan usein ns. *bottom-up*-lähestymiseksi, jossa oppimisen lähtökohtana ja myös sen ennakkoehtona ovat todelliset kielen ilmaukset, joita kielenkäyttäjä kohtaa ja joiden havaitut frekvenssit ja sisäinen variaatio ovat keskeisimpiä oppimista ja rakenteiden skemaattistumista määrittäviä tekijöitä (Taylor 2002: 592).

Kielen rakenteen oppimista tarkasteltaessa käyttöpohjainen viitekehys on osoittautunut monin tavoin hyvin perustelluksi, sillä sen avulla on pystytty uskottavasti kuvaamaan oppimista edeltävän syötöksen ja syötöksen synnyttämän oppimisen välistä suhdetta sekä kielenoppimisprosessissa kehkeytyvän abstraktinkin kielitaidon aineistolähtöisyyttä. (Ks. esim. Bates–MacWhinney 1987; Tomasello 1992; Lieven ym 1997.) Esimerkiksi Brian MacWhinney kuvaa käyttöpohjaisesti rakenteen ja rakenteiden kehkeytymistä toiston (engl. *rote*), yhdistämisen (engl. *combination*) ja analogian (engl. *analogy*) määrittämänä ilmiönä (MacWhinney 2001: 457–459). Oppimista voi tapahtua vasta silloin, kun voidaan tunnistaa uuden syötöksen osia ja yhdistää ne parhaiten uutta syötöstä vastaaviin vanhemman syötöksen osiin. Näin käytetystä kielestä muodostuu vähitellen säännöllisyyksiä, joita oppija havaitsee yhä paremmin ja hienojakoisemmin vastaanotetun syötöksen määrän kasvaessa. (Mts. 464.) Kyky analogiseen havainnointiin mahdollistaa edelleen tällaisten monimutkaisten kokonaisuuksien välisten rakenteellisten samankaltaisuuksien tunnistamisen ja täten

rakenteellisen oppimisen. Näin syntyvien abstraktioiden pohja on kuitenkin nimenomaan aiemmissa havainnoissa ja näiden aiempien havaintojen välisissä luonnollisessa käytössä havaituissa samankaltaisuuksissa sekä muodon että merkityksen osalta (ks. Tomasello 1992: 238–257).

2.1.2. Konstruktiot kielen rakenteen kuvaajina ja opittavina ilmiöinä

Käyttöpohjaisessa viitekehyksessä kielen rakennetta ja erilaisia kompositionaalisia kokonaisuuksia tarkastellaan usein konstruktiokieliopin näkökulmasta. Termiä konstruktio on käytetty jo pitkään kuvaamaan hyvin erilaisia kielen rakenneosia yksittäisistä sananmuodoista aina leksikaalisesti täysin rajaamattomiin lausekonstruktioihin. Konstruktiokäsitteen merkitys ei ole välttämättä aina yksiselitteinen, ja eri tutkijat ovatkin tarkoittaneet termillä hieman eri asioita. Lari Kotilainen (2007) tiivistää konstruktioajattelulle yhteiset keskeiset teesit kolmeen kohtaan:

1. Kielioppi koostuu konstruktioista.
 2. Konstruktioissa yhdistyy tieto muodosta ja merkityksestä.
 3. Kielioppi on järjestäytynyt.
- (Kotilainen 2007: 19–21)

Perustavanlaatuisina yhdistävinä tekijöinä voidaan siis pitää sitä, että konstruktiot ovat kielen muotopiirteiden ja niiden funktioiden välisiä vastaavuuksia ja että nämä muoto–merkitys-parit ovat kielijärjestelmän perusyksiköitä (ks. esim. Fillmore 1985; Fillmore, Kay & O'Connor 1988; Kay 1995; Goldberg 1995, 2006; Croft 2001). Keskeistä käsitykselle on lisäksi se, että konstruktioiden järjestelmä on monikerroksinen ja hierarkkinen jatkumo eli samalla järjestelmällä voidaan kuvata kielen eri tasojen ilmiöitä sanastosta ja morfologiasta syntaksiin niin, että konstruktiot jäsenyvät suhteessa toisiinsa analogisesti eli funktionaalisen rakenteellisen samankaltaisuuden kautta² (esim. Kay 1995: 177–179; Goldberg 1995: 108–110; Goldberg 2006: 13–14). Alla olevien esimerkkien 1–6 rakenteet kuvaavat konstruktioiden keskinäistä erilaisuutta ja sitä, että kieliopin eri tasot koostuvat nimenomaan konstruktioista. Esimerkit ovat Kotilaiselta (2007: 17; ks. myös Croft 2001: 17; Goldberg 2006: 5). Käyttöpohjaisen ajattelun kannalta on syytä huomioida se, että konstruktiokuvauksissa yksittäisen konstruktion eri osat voidaan määritellä keskenään eri abstraktiotasolla (ks. esimerkin 2 omistusrakenne, joka koostuu *olla*-verbistä yksikön

² Kuten edellä mainitsen, konstruktiokieliopissa tätä analogista suhdetta kutsutaan usein ominaisuuksien periytyvyydeksi (engl. *inheritance*). Itkonen kuitenkin huomauttaa aiheellisesti, että kyseessä on nimenomaan eri abstraktiotasoilla esiintyvä rakenteellinen samankaltaisuus eli analogia (Itkonen Esa 2011: 602; analogian käsitteestä konstruktiokieliopissa ks. myös Hamunen 2015).

kolmannessa persoonassa ja kahdesta leksikaalisesti rajoittamattomasta substantiivilausekkeesta, joiden on kuitenkin oltava tietyssä morfologisessa muodossa).

1) lausetyyppi (transitiiviskeema)	$NP_{\text{NOM}} + V + NP_{\text{PART/GEN/NOM}}$
2) lausetyyppi (omistus rakenne)	$NP_{\text{ADE}} + \text{y.3.p. verbi } olla + NP_{\text{NOM}}$
3) idiomi	<i>mennä vipuun</i>
4) yhdyssana	<i>kalakukko</i>
5) sana	<i>liekki</i>
6) morfeemi (ei vapaa) (Kotilainen 2007: 17)	elatiivin päätte (- <i>sta</i>)

Konstruktikieliopin erilaisia sovelluksia erottaa toisistaan selvimmin se, korostavatko ne kieliopin generativistisuutta, kuten esimerkiksi Paul Kay (1995: 172), vai sen kognitiivista luonnetta, kuten esimerkiksi William Croft (2001) ja Goldberg (1995; 2006). Kognitiivista kielentutkimusta voidaan käyttää laajana kattoterminä sellaisille kielitieteen suuntauksille, joiden mukaan kieli ei ole muista kognitiivisista kyvyistä erillään oleva järjestelmä, joissa kielioppi nähdään tapana käsitteistää kokemuksia ja joissa kielijärjestelmän katsotaan syntyvän todellisesta kielenkäytöstä (Croft & Cruse 2004: 1). Käyttöpohjaiset mallit ovat siis luonteeltaan nimenomaan kognitiivisia (ks. Langacker 1990: luku 10), ja tämän tutkimuksen konstruktioikäisy nojaakin ennen kaikkea kognitiiviseen kielentutkimukseen orientoituneisiin tutkimuksiin. Kognitiivistenkin konstruktikieliopin sovellusten välillä on jonkin verran eroa siinä, onko konstruktion yhtenä määrittelykriteerinä sellainen konstruktioimerkitys, joka ei ole suoraan pääteltävissä konstruktion osien merkityksistä. Nojaan tässä työssä Goldbergin myöhempään määritelmään (Goldberg 2006; vrt. Goldberg 1995: 4), jonka mukaan konstruktioilta ei välttämättä vaadita erillistä konstruktioimerkitystä, mikäli saman muodon ja merkityksen yhdistyminen on kyllin frekventtiä:

"ALL LEVELS OF GRAMMATICAL ANALYSIS INVOLVE CONSTRUCTIONS: LEARNED PAIRINGS OF FORM WITH SEMANTIC OR DISCOURSE FUNCTION, including morphemes or words, idioms, partially lexically filled and fully general phrasal patterns. [...] Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency".
(Goldberg 2006: 5.)

Frekvenssin nostaminen osaksi kielellisen muoto-merkitys-parin määritelmää on keskeistä käyttöpohjaiselle kielikäsitteilylle. Käytetyn kielen määrällisten ominaisuuksien ja niistä kumpuavien tilastollisten vihjeiden ja säännönmukaisuuksien tunnistamisen on ylipäätään

havaittu olevan keskeisiä mekanismeja myös kielen oppimisessa (ks. esim. Elman 1990; Marcus et. al. 1999; Saffran 2001; Saffran & Wilson 2003). Perusajatuksena käyttöpohjaisessa rakenteen oppimisessa on prosessin käynnistyminen ja eteneminen yksikkökohtaisesti, todellisten ilmausten perusteella. Rakenteesta ei siis opita sääntöjä, vaan oppija kohtaa todellisen kielen ilmauksia, joiden väliset rakenteelliset samankaltaisuudet ja erot mahdollistavat esimerkkejä noudattavien yleistysten tekemisen analogisesti myös uusia ilmauksia muodostettaessa.

Elisabeth Bates ja Judith Goodman antavat hyvin konkreettisen esimerkin kielen käyttöpohjaisesta luonteesta osoittaessaan, että ensikieltä opittaessa oppijan sanaston koko ja syntaktiset taidot korreloivat voimakkaasti keskenään (Bates & Goodman 1997). Vastaavasti Klaus Laalon (2011) tutkimus suomenkielisten lasten verbintaivutuksen kehityksestä kuvaa pitkälti tällaista prosessia, samoin Michael Tomasellon (1992) tutkimus englanninkielisten lasten verbikonstruktioiden oppimisesta. Laalon mukaan suomen verbintaivutuksen alkuvaiheen oppiminen keskittyy muutamaa taajaan esiintyvään verbilekseemiin (Laalo 2011: 108–109) ja etenkin näiden kahteen muotoon: yksikön kolmannen persoonan indikatiivin preesensiin (esim. *katsoo*) ja yksikön toisen persoonan imperatiiviin (esim. *katso*) (mts. 111). Vastaavasti Tomasellon mukaan englannin verbi-ilmaukset opitaan yleensä verbikohtaisesti niin, että tietty argumenttirakenne esiintyy aluksi vain yhden verbilekseemin kanssa, vaikka ilmauksen muut osat voivatkin vaihdella (Tomasello käyttää termiä *verb island*, Tomasello 1992). Molemmissa tutkimuksissa nämä alkuvaiheen lekseemit olivat opittavan kielen taajaan esiintyviä sanoja. Goldberg ym. (2004) osoittavatkin osin Tomasellon tutkimukseen nivelyten, että nämä oppimisen alkuvaiheen verbikonstruktioiden verbit ovat niin esiintymiseltään kuin merkitykseltäänkin prototyyppisiä sille rakenteelle, jossa ne esiintyvät. Kaikki nämä tutkimukset osoittavat niin ikään, että vähitellen kokemuksen ja syötöksen määrän lisääntyessä rakenteet kuitenkin kehkeytyvät abstraktimmiksi ja analogiamuodosteiden määrä lisääntyy.

Edellä kuvatun kaltaista kielen rakenteiden kehkeytymistä ja niiden produktiivisuutta tarkastellaan käyttöpohjaisesta näkökulmasta usein havainnoimalla tarkastelunalaisen konstruktion esiintymäfrekvenssiä (engl. *token frequency*) ja tyyppifrekvenssiä (engl. *type frequency*), sekä niiden keskinäistä suhdetta ja sen vaikutusta syntaksiin. Esiintymäfrekvenssillä tarkoitetaan tyypillisesti sitä, kuinka monta kertaa jokin tietty konstruktio esiintyy jossakin tietyssä aineistossa. Tyyppifrekvenssi puolestaan kertoo sen,

miten paljon sisäistä vaihtelua konstruktiossa voi esiintyä eli esimerkiksi kuinka monen eri verbin kanssa jotakin tiettyä argumenttirakennekonstruktioita voidaan käyttää. (Bybee & Thompson 1997.) Niinpä esimerkiksi suomen kielen transitiivikonstruktion (esim. 7) ja eksistentiaalikonstruktion (esim. 8) esiintymäfrekvenssi kertoo, kuinka monta kumpaakin konstruktioita jossakin aineistossa esiintyy, kun taas konstruktioiden tyyppifrekvenssi voi kertoa esimerkiksi sen, kuinka monta eri verbiä konstruktion predikaattiasemassa esiintyy.

7) Jaakko syö leipää.

8) Tallissa on auto.

Molemmat konstruktiot ovat frekventtejä eli niiden esiintymäfrekvenssi on korkea. Transitiivikonstruktion toisen jäsenen – predikaatin – tyyppifrekvenssi on kuitenkin huomattavasti eksistentiaalikonstruktioita korkeampi, sillä eksistentiaalikonstruktiossa verbit ovat lähes aina yksikön kolmannessa persoonassa ja niiden on oltava staattisia, intransitiivisia ja olemassa olemista tai siihen liittyviä prosesseja kuvaavia (VISK §893). Aiemmassa tutkimuksessa erityisesti termillä *token frequency* viitataan joko eksplisiittisesti tai implisiittisesti sanojen esiintymismääriin (ks. Bybee & Thompson 1997: 378), mistä syystä suomenkielisenä terminä käytössä on usein sanefrekvenssi (ks. esim. Jantunen 2004: 98). Nähdäkseni konstruktiokieliopin viitekeh്യksessä termi sanefrekvenssi ei kuitenkaan vastaa ilmiön koko kirjoa. Se korostaa sanojen merkitystä muunlaisten kielenpiirteiden kustannuksella ja pyrkii luokittelemaan ilmiötä, joka on luonteeltaan selvästi jatkumomainen. Konstruktiokieliopin näkökulmasta skemaattisuudeltaan erilaisia konstruktioita ei eroteta periaatteellisella tasolla toisistaan, ja esiintymä- ja tyyppifrekvenssin suhdetta tarkasteltaessa kumpikaan frekvenssi ei välttämättä kuvaa mitään sellaista kielenpiirrettä, joka olisi luonteeltaan yksiselitteisen leksikaalinen. Esimerkin 8 eksistentiaalikonstruktiossa voi ensimmäisenä jäsenenä olla laaja kirjo substantiiveja – tyyppifrekvenssi on hyvin korkea – kun taas morfologisina muotoina siihen käyvät lähinnä esimerkissäkin näkyvä inessiivi ja adessiivi (esim. *pöydällä on kirja*), eli tyyppifrekvenssi on morfologisten muotojen osalta rajatumpi. Käytänkin tässä väitöskirjassa termiä esiintymäfrekvenssi kuvaamaan jonkin konstruktion esiintymien määrää ja termiä tyyppifrekvenssi kuvaamaan jossakin konstruktiossa havaittavan variaation määrää.

Konstruktioiden produktiivisuutta voidaan tarkastella vertaamalla niiden esiintymäfrekvenssin ja tyyppifrekvenssin välistä suhdetta, ja tämän suhteen katsotaankin kuvaavan konstruktioiden skemaattisuutta. Lisäksi näiden frekvenssien välinen suhde

vaikuttaa siihen, miten kielijärjestelmä muuttuu. (Ks. esim. Bybee 1985; Bybee & Thompson 1997.) Oppimisen kannalta on keskeistä, että prosessi perustuu todelliseen kielenkäyttöön. Näin ollen konstruktiokohtaisen oppimisen voi katsoa alkavan silloin, kun konstruktio esiintyy ensi kerran tunnistettavassa muodossa. Tyypifrekvenssin eli konstruktiossa havaittavan variaation kasvu puolestaan kuvaa oppimisen vähittäistä kehitystä. (Martin ym. 2010, ks. myös 3.3.) Mitä korkeampi on jonkin ilmauksen esiintymäfrekvenssi, sitä todennäköisemmin se myös vakiintuu kieleen sellaisenaan. Näin esimerkiksi kielen kaikkein frekventeimmillä sanoilla on absoluuttisesti korkea esiintymäfrekvenssi ja monilla idiomikonstruktioilla on suhteellisesti korkea esiintymäfrekvenssi eli idiomien osat esiintyvät kokonaisfrekvenssiinsä nähden usein samassa muodossa ja osana samanlaista fraasia. Vastaavasti produktiivisilla rakenteilla on leksikon tasolla tarkasteltaessa korkea tyypifrekvenssi, sillä rakenteen samassa roolissa voi esiintyä useita eri lekseemejä. Vähitellen oppijalle hahmottuu kielen tyypillisiä rakenteita, jotka muodostuvat erilaisista elementeistä ja näitä elementtejä koskevista rajoituksista ja preferensseistä.

Kuten edellä totesin, tässä tutkimuksessa jonkin kielenpiirteiden joukon tulkinta konstruktioksi ei edellytä sitä, että joukolla olisi oma konstruktiomerkitys, jota ei voida johtaa sen osien merkityksestä. Puhuessani tässä työssä konstruktioista tarkoitan siis vakiintunutta muodon ja merkityksen yhdistelmää riippumatta siitä, onko sen yhteismerkitys tulkittavissa suoraan niistä osista, jotka muodostavat kyseisen konstruktion. Muodon ja merkityksen välisen yhtenevyyden on kuitenkin toistuttava. Tämä on relevanttia jo frekvenssiin ja toistuvuuteen perustuvan kielikäsitteilyn kannalta, sillä konstruktioiden oppimisen katsotaan perustuvan analogisiin havaintoihin eli tässä tapauksessa rakenteellisesti samankaltaisten kielenpiirteiden joukkojen toistumiseen keskenään samankaltaisessa merkityksessä. Itkonen huomauttaa lisäksi, että tyypillisesti myös merkitykseltään ei-kompositionaaliset konstruktiot määrittellään yleensä nimenomaan osiensa perustella, ja tässä mielessä koko konstruktion käsite on määrittelyn tasolla luonteeltaan kompositionaalinen (Itkonen, Esa 2011). En pyri ottamaan kantaa tarkastelemieni konstruktioiden perimmäiseen kompositionaalisuuteen, mutta uskon kielenpiirteiden toistuvan yhteisesiintymisen osoittavan jonkinlaista rakennemaisuutta silloin, kun rakenteen osien yhteisesiintyminen on selvästi sattumanvaraista yhteisesiintymistä todennäköisempää. Rakenteen ja konstruktioiden tarkastelu eri osatekijöiden ja näiden yhteisesiintymisten frekvensseihin perustuen tekeekin koko tämän tutkimuksen tarkastelusta jo lähtökohtaisesti luonteeltaan kompositionaalista.

2.1.3. Fraseologisten yksiköiden ja konstruktioiden suhde

Konstruktiokieliopin piirissä tehtävän tutkimuksen tapaan myös fraseologisessa tutkimuksessa keskiössä on todellinen kielenkäyttö. Käyttöön pohjautuvan perusluonteen lisäksi fraseologisen tutkimuksen ja konstruktio-ajattelun väliset yhtäläisyydet ovat niin selviä, että niiden pitäminen erillään tuntuu keinotekoiselta. Fraseologian ja fraseologisten yksiköiden käsitteiden merkitys vaihtelee konstruktioiden tapaan tarkastelunäkökulmasta riippuen. Keskeisenä erona näiden tarkastelumallien välillä on se, että fraseologinen tutkimus on kiinnostunut ennen kaikkea yhtä leksikaalista elementtiä laajemmista kielenyksiköiden joukoista. Yhteistä konstruktioiden ja fraseologisten yksiköiden tutkimuksen välillä sen sijaan on kiinnostus kielenyksiköiden tyypilliseen sisäiseen ja kontekstuaaliseen käyttöön, merkitykseen ja siinä esiintyvään vaihteluun. Stefan Gries (2008) nostaa esiin kuusi sellaista fraseologisen yksikön käsitettä määrittävää parametriä, joihin fraseologisen tutkimuksen olisi hänen mukaansa vastattava, jotta saatavia tuloksia voidaan suhteuttaa esimerkiksi konstruktio-ajattelun mukaiseen tutkimukseen:

1. fraseologiseen yksikköön kuuluvien elementtien luonne
 2. fraseologiseen yksikköön kuuluvien elementtien määrä
 3. fraseologisen yksikön havaintojen kynnysmäärä
 4. fraseologisen yksikön elementtien välinen hyväksyttävä etäisyys
 5. fraseologiseen yksikköön kuuluvien elementtien leksikaalisen ja syntaktisen joustavuuden aste
 6. semanttisen yhteneväisyyden ja ei-kompositionaalisuuden rooli fraseologisen yksikön määritelmässä
- (Gries 2008: 4–6)

Näiden parametrien valossa Griesin oma fraseologinen asemoituminen on hyvin konstruktiomainen. Hän näkee fraseologisen yksikön kriteereinä sen, että vähintään yksi yksikön elementeistä on määriteltävä jossain määrin leksikaalisesti, että yksikön muodostavia elementtejä on oltava vähintään kaksi, että yhteisesiintymien havaittu frekvenssi ylittää sattuman eli odotetun frekvenssin ja että fraseologisen yksikön on oltava semanttisesti yhteneväinen mutta ei välttämättä ei-kompositionaalinen (Gries 2008: 4–6).

Keskeisin ero Griesin esittämän fraseologisten yksiköiden määritelmän ja edellä kuvatun konstruktio-näkökulman välillä on se, että määritelmä sulkee pois transitiivikonstruktion kaltaiset, leksikaalisesti täysin spesifioimattomat argumenttirakennekonstruktiot. Gries keskittyy lisäksi vain kompositionaalisuuteen eli hän rajaa tarkastelun ulkopuolelle sellaiset tapaukset, joissa on vain yksi elementti. Lisäksi suomen kaltaisten, rikkaan taivutusjärjestelmän avulla jäsentyvien kielten asema Griesin hahmotelmassa jää epäselväksi. Lienee tulkinnanvaraista, ovatko vaikkapa nominien eri sijamuodot tai verbien

persoonapäätteet leksikaalista ainesta vai eivät. Kirjaimellisesti tulkiten niitä ei voida lukea leksikaalisesti määriteltäväksi osaksi mutta toisaalta esimerkiksi indoeurooppalaisten kielten adpositiot nähdään leksikaalisena aineksena. Lisäksi Griesin malli ei ota suoraan kantaa, mitä leksikaalinen rajaus tarkkaan ottaen tarkoittaa. Kielessä on lopulta melko vähän konstruktioita tai fraseologisia yksiköitä, jotka ovat edes semanttisesti täysin rajaamattomia, ja esimerkiksi suomen transitiivikonstruktioita voi tarkastella jatkumona sen suhteen, missä määrin se on transitiivinen (VISK §892). Näin ajatellen on epätodennäköistä mutta mahdollista, että suomen transitiivikonstruktion subjekti-NP:nä esiintyy jokin merkitykseltään eloton ja hyvin epäagentiivinen sana (ks. esimerkki 9). Toisaalta kielessä on runsaasti 'ei ole kaikki x:t y:ssä' -konstruktion kaltaisia (Kortelainen 2011) selkeitä idiomikonstruktioita, joissa voi kuitenkin olla runsaasti leksikaalista variaatiota. Lisäksi leksikaalisesti hyvinkin rajatuissa idiomimaisissa konstruktioissa voi niin ikään analogisesti soveltaa myös konstruktiolle epätyypillisiä semanttisesti tai fonologis-morfologisesti samankaltaisia sanoja siten, että konstruktiomerkitys kuitenkin säilyy. Näin esimerkiksi niin sanottuun myöntökieltokonstruktioon (Miestamo 2004; Kotilainen 2012) voi soveltaa myös muita sanoja kuin kiro sanoja, vaikka ne ovatkin konstruktiolle odotuksenmukaisimpia (ks. esim. 10).

9) Jaakko syö leipää ~ Patsas syö leipää.

10) Vitut mä mitään tee ~ Ketut mä mitään tee.

Fraseologisessa tutkimuksessa voidaan erottaa perinteinen fraseologinen lähestymistapa (engl. *phraseological approach*) ja frekvenssiperustainen lähestymistapa (engl. *frequency-based approach*) (Granger & Paquot 2008: 28–29; merkitys- ja toistuvuussidonnaisia määritelmiä kuvaavista termeistä ks. Nesselhauf 2004: 11–12). Sylviane Granger ja Magali Paquot suosittelevat näiden lähestymistapojen erottamista ennen kaikkea siitä syystä, että sanojen tai kielellisten elementtien frekventti yhteisesiintyminen ei vielä itsessään kerro niiden muodostavan fraseologista merkitysyksikköä (Granger & Paquot 2008: 41), vaan frekvenssiperustaisesti havaitun yhdessä esiintyvien kielenpiirteiden joukon osoittaminen fraseologiseksi yksiköksi edellyttää laadullista analyysia ja siinä havaittavaa merkityksen yhdenmukaisuutta. Vastaavasti Gries jättää edellä esitettyihin parametreihinsa vedoten pois sellaiset jo yksinään frekventtien elementtien yhteisesiintymät, jotka eivät muodosta semanttista kokonaisuutta (Gries 2008: 8). Ajatus tuntuu luontevalta myös aiemmin esittämäni konstruktiokieliopin suuntauksen kannalta, sillä konstruktiolta vaaditaan toistuvuuden lisäksi myös funktionaalista yhdenmukaisuutta eli muoto–merkitys-parin konventionaalitumista (Goldberg 2006: 3).

Granger ja Paquot korostavat fraseologisen tutkimuksen liikkuvan samanaikaisesti morfologian, syntaksin, semantiikan ja diskurssin rajapinnoilla (Granger–Paquot 2008: 28). Keskiössä on tyypillisesti vapaiden yhdistelmien ja kivettyneiden idiomien välinen jatkumo (ks. esim. Cowie 1981; Sinclair 1991: 110) ja jonkin monisanaisen yksikön sijoittuminen tälle jatkumolle. Kielen monisanaiset yksiköt siis sijoittuvat tälle jatkumolle sen mukaan, minkälaista ja kuinka laajaa leksikaalista tai muotopiirteellistä vaihtelua yksikössä yleensä esiintyy tai ylipäättään voi esiintyä. John Sinclair kuvaa tätä todennäköisen ja mahdollisen vaihtelun välistä suhdetta idiomiperiaatteen (engl. *idiom principle*) ja vapaan valinnan periaatteen (engl. *open choice principle*) välisenä jatkumona. Näin hän kuvaa sitä, että monissa kielen ilmauksissa on periaatteessa todella paljon vaihtoehtoja, joista jotkut ovat kuitenkin huomattavasti odotuksenmukaisempia eli idiomaattisempia kuin toiset. Idiomaattisuuteen tai odotuksenmukaisuuteen liittyvät seikat voivat näkyä monin tavoin kieliopillisten rajoitusten lisäksi esimerkiksi erilaisina esiintymisrajoituksina tai -preferensseinä. (Sinclair 1991: 109–110.)

Vastaavaa ilmiötä on tarkasteltu myös konstruktiokieliopin näkökulmasta, ja usein konstruktiomerkitys ja sen spesifisyys korreloivat vahvasti tarkasteltavan konstruktion esiintymä- ja tyypifrekvenssin suhteen kanssa. Esimerkiksi Anatoli Stefanowitsch ja Stefan Gries (2003) sekä Adele Goldberg ym. (2004) toteavat, että monissa lausetason verbikonstruktioissa pieni määrä verbileksemejä esiintyy valtaosassa kaikista tietyn konstruktion esiintymistä. Useiden verbikonstruktioiden määrällisesti tyypillisimmän verbileksemin merkitys on myös lähinnä sen konstruktiomerkitystä. Voidaankin ajatella, että tietyn rakenteen frekventein esiintymismuoto vaikuttaa myös kaikkien muiden saman rakenteen esiintymien merkitykseen ja konstruktion kokonaismerkityksen oppimiseen. (Goldberg ym. 2004: 297–299.)

Kielenpiirteiden esiintymisen ja yhteisesiintymisen toistuvuus luonnollisessa kielenkäytössä on perusteltu lähtökohta, kun kielen oppimista tutkitaan aineistovetoisesti käyttöpohjaisen mallin näkökulmasta. Lähestymistapa mahdollistaa luontevasti myös tyypillisten oppijankieltä ja ensikieltä toisistaan erottavien piirteiden havainnoinnin ja eron luonteen arvioinnin. Lähestymistapa on hyvin yleinen korpuskielitetieteellisessä tutkimuksessa ylipäättään, ja korpustutkimuksen parissa suosittu, niin kutsuttu firthiläinen kieliteoria perustuu pitkälti nimenomaan luonnollisilla kieliaineistoilla tehtävään kielentutkimukseen (Firth 1968; ks. myös Sinclair 1966; Palmer 1976; Stubbs 1996; Tognini-Bonelli 1996, 2001).

Lähtökohtana on kontekstuaaliseksi semantiikaksi kutsuttu ajatus siitä, että kielessä merkitys muodostuu kontekstilähtöisesti kulloinkin tarkasteltavana olevan kielenpiirteen suhteessa yhtäältä tekstuaalisen lähikontekstin ja toisaalta laajemman tilannekontekstin kanssa (yhteenvetona ks. esim. Stubbs 1996). Jarmo Jantunen (2004) tiivistää kontekstuaalisen semantiikan pyrkimykset seuraavasti:

”Kontekstuaalinen semantiikka ei pyri etsimään abstrakteja käsitteitä, jotka kielenkäyttäjät jakavat, vaan se keskittyy tutkimaan konkreettisia toistuvia käyttöyhteyksiä ja näin kuvaamaan niiden usuaalista merkitystä. Tutkimuksen keskiössä on erityisesti ilmausten syntagmaattisten ja paradigmaattisten ulottuvuuksien kohtauspiste eli ne säännöt tai preferenssit, joiden perusteella ilmaus valitaan muiden ilmausten joukosta (paradigmaattinen suhde) muiden ilmausten kanssa lineaariseen jatkumoon (syntagmaattinen suhde).”
(Jantunen 2004: 8.)

Konstruktiokieliopin yleisesti hyödyntämä kehyssemanttinen merkitysnäkemys on nähdäkseni hyvin samankaltainen kuin edellä kuvattu kontekstuaalinen merkitysnäkemys: kielelliset merkitykset eivät ole koskaan olemassa itsessään vaan ne suhteutuvat aina johonkin kielenkäyttäjien itsensä tuottamiin kielenulkoisiin eli kontekstuaalisiin semanttisiin kehyksiin (ks. esim. Fillmore 1975, 1977; Goldberg 1995; Västi 2012: 49–53). George Lakoff (1987) antaa hyvän esimerkin kehysten (Lakofilla *idealized cognitive model*) kontekstuaalisesta merkityksestä kuvatessaan *mother*-sanana (’äiti’) erilaisia kontekstuaalisia merkitystulkintoja:

- 11) I was adopted, and I don’t know who my real mother is.
 - 12) I am not nurturant person, so I don’t think I could ever be a real mother to any child.
 - 13) My real mother died when I was an embryo.
 - 14) I had a genetic mother who contributed the egg that was planted in the womb of my real mother.
- (Lakoff 1987: 75.)

Esimerkeissä *mother*-sanana tulkinta riippuu vahvasti siitä, mitä edeltävistä lauseista tarkastellaan. Kunkin sanan kielellinen esiintymiskonteksti spesifioi sen, minkälainen *mother*-sanana merkityksen suhde kielenulkoiseen maailmaan on kussakin tapauksessa etusijalla. Niinpä esimerkki 11 korostaa äidin merkitystä synnyttäjänä, esimerkki 12 huolenpitäjänä, kun taas esimerkit 13 ja 14 kuvaavat erilaisia näkemyksiä sanan suhteesta perintötekijöihin ja asuinympäristöön. Kaikki nämä tulkinnat perustuvat sanan lähikontekstiin ja sen suhteeseen kielenulkoiseen maailmaan. Kaikki kuvaavat osaltaan sanan *mother* merkitystä, ja toisaalta yleisimmät käyttök kontekstit vaikuttavat merkitykseen harvinaisia enemmän.

Edellä kuvatun kaltaista kontekstisidonnaista tulkinnanvaraisuutta esiintyy niin leksikaalisesti spesifeissä kuin skemaattisemmissakin konstruktioissa. Kontekstuaalisen semantiikan tuottamat merkityskuvaukset ovat luonteeltaan vahvasti käyttöpohjaisia, sillä tarkasteltavien kielenpiirteiden merkityksiä ja lähimerkityksisten tapausten eroa jäsennetään yleensä niiden tyypillisen käytön kannalta. Tyypillinen käyttö eli tarkasteltavan kielenpiirteen yleinen esiintymisympäristö koostuu myötäesiintymistä eli sellaisista sanoista, kielellisistä muodoista tai merkityksistä, joiden esiintyminen tarkasteltavan kielenpiirteen yhteydessä on yleistä. Ajatus vastaa edellä kuvattua idiomiperiaatetta, jonka mukaan ilmausten käyttökontekstit ja kielenpiirteiden yhteisesiintymiset sijoittuvat vapaasti valittavien ja täysin rajoitettujen käyttöympäristöjen väliselle jatkumolle. (Jantunen 2004: 15.) Vastaavasti sekä idiomiperiaate että kontekstuaalinen semantiikka vastaavat periaatteellisella tasolla edellä kuvattua ajatusta siitä, että kielen rakenteet ja niiden merkitys muotoutuvat rakenteen esiintymäfrekvenssin ja rakenteen sisäisen ja sen tyypillistä ympäristöä kuvaavan tyypifrekvenssin sekä näissä havaittavien tyypillisyyksien ohjaamana.

Todellisten kielellisten ilmausten ja kielenulkoisten merkitysten välisten parien korostaminen kuvastaa yhtäältä vahvaa yhteyttä fraseologisen tutkimuksen ja konstruktiokieliopin välillä ja toisaalta näiden molempien läheisyyttä todelliseen kielenkäyttöön ja sitä kautta käyttöpohjaiseen kielikäsitukseen. Käytän tässä väitöskirjassani termiä konstruktio kuvaamaan hyvin monenlaisia kieliopin yksiköitä yksittäisistä morfologisista muodoista monisanaisiin lauserakenteisiin (ks. *S2-avainrakenteet*, *Siirtovaikutus-avainrakenteet*, *Pitkittäismuutokset* ja *Mahdollisuuden ilmaukset*), ja keskeisinä kriteereinä ovat määrällinen toistuvuus ja merkityksen koherenssi. Kutsun näitä kielen yksiköitä konstruktioiksi, ja tutkimukseni nivoutuukin sekä teoreettisesti että terminologisesti osaksi kognitiivista konstruktiokielioppia ja sitä kautta käyttöpohjaisiin teorioihin perustuvien tutkimusten joukkoa. En kuitenkaan näe mitään periaatteellista eroa valitsemani terminologis-teoreettisen ratkaisun ja fraseologisen tutkimuksen terminologian ja teoreettisten perusolettamusten välillä. Niinpä sovellan tutkimuksissani konstruktiokieliopin lisäksi runsaasti myös aiempaa fraseologista tutkimusta ja fraseologisen tutkimuksen metodologiaa sekä tarvittaessa myös fraseologisessa tutkimuksessa käytettyä terminologiaa.

2.1.4. Kohteena konstruktiot – lähtökohtana toistuvat mallit

Olen edellä kuvannut sekä kognitiivisesti orientoituneen konstruktiokieliopin että fraseologisen tutkimuksen näkemyksiä kielen rakenteesta ja kieliopista. Molemmille

lähestymistavoille on yhteistä se, että niissä korostetaan kielen toisteisuutta sen rakennetta määrittävänä mekanismina. Lisäksi molemmat lähestymistavat edellyttävät yleensä sitä, että toisiaan muistuttavien kielen jaksojen täytyy muistuttaa toisiaan myös merkitykseltään, jotta ne voidaan käsittää saman rakenneyksikön edustajiksi. Toisin sanoen kaikki toistuvasti yhdessä esiintyvät kielenpiirteet eivät ole konstruktioita tai fraseologisia merkitysyksikköjä. Käytän termiä toistuva malli (engl. *recurring pattern*) tietynlaisena kattoterminä kuvaamaan kaikenlaisia toistuvia kielellisiä asetelmia ja kielenpiirteiden esiintymistä yhdessä. Toistuvat mallit voivat olla joko itsessään konstruktioita, ne voivat kuvata osaa jostakin konstruktioista tai ne voivat sisältää useita konstruktioita osia – ainoana kriteerinä on samankaltaisten kielenpiirteiden joukon toistuva yhteisesiintyminen. Toistuvilta malleilta ei siis vaadita samankaltaisen merkityksen toistumista, mistä syystä yksi toistuva malli saattaa kuvata esimerkiksi useita eri konstruktioita tai usean eri konstruktion osia – kuten suomen *n*-sija, joka saattaa esiintyä niin määritteenä tai omistajan ilmaisijana genetiivikonstruktiossa (*Jaakon koira*) kuin totaaliobjektina transitiivikonstruktiossakin (*näin eilen Jaakon*). Toisaalta toistuvat mallit saattavat kattaa osia useista eri konstruktioista ja näin kuvata niiden tyypillistä käyttöä. Niinpä esimerkiksi partitiivista, konjunktiosta ja nominatiivista koostuva toistuva malli voi kuvata esimerkiksi virkkeensisäisten lauseiden alistussuhdetta, kuten esimerkin 15 *satua että lapsi* tai rinnastussuhdetta, kuten esimerkin 16 *satua ja lapsi*:

15) Äiti lukee illalla satua, että lapsi nukahtaisi.

16) Äiti lukee illalla satua ja lapsi kuuntelee.

Etsin ja tutkin tässä väitöskirjatutkimuksessa toistuvien mallien avulla konstruktioita, joiden käyttö erottaa *S1*- ja *S2*-kielenkäyttäjiä (ks. *S2-avainrakenteet*) tai kielitaustaltaan erilaisia *S2*-kielenkäyttäjiä (ks. *Siirtovaikutus-avainrakenteet*) toisistaan tai joiden käyttö muuttuu pitkittäistarkastelussa (ks. *Pitkittäismuutokset*). Vaikka havaitut konstruktiot voivat koostua keskenään hyvin erilaisista kielenpiirteistä – kuten *esimerkiksi*-sanamuoto tai sanajärjestykseltään käänteinen predikatiivikonstruktio – voidaan erot nähdäkseni paikantaa ja joiltakin osin myös selittää nimenomaan toistuvien mallien ja niissä havaittujen frekvenssierojen avulla.

Kuvaan seuraavaksi niitä teoreettisia käsitteitä, joiden avulla tarkastelen aineistossa havaittavia toistuvia malleja. Käytännössä kaikki käsitteet nivELYTÄVÄT jonkin kielenpiirteen sisäiseen variaatioon (kuten jonkin lekseemin eri morfologiset muodot, jonkin morfologisen muodon eri syntaktiset funktiot tai syntaktisen funktion eri lekseemit) ja niiden

esiintymisympäristössä ilmenevään variaatioon (engl. *item and environment*, ks. Francis 1993). Tarkastelen toistuvia malleja ja jäsenen kielenpiirteiden kielellistä lähikontekstia neljän kielenpiirteiden yhteisesiintymistä kuvaavan käsitteen avulla. n-grammit (engl. *n-gram*), kollokaatiot (engl. *collocation*), kolligaatiot (engl. *colligation*) ja kollostruktiot (engl. *collostruction*) kuvaavat erilaisia tapoja lähestyä kielenpiirteiden yhteisesiintymistä. Näiden käsitteiden avulla voidaan tarkastella jonkin kielenpiirteen kielellistä lähikontekstia eli kotekstia. Erotankin kielellisen lähikontekstin yleisemmästä konteksti-termistä, jolla viitataan ennen kaikkea laajempaan tilanteeseen ja kielenulkoiseen käyttöympäristöön. Kaikki nyt esiteltävät kielellistä lähikontekstia jäsentävät käsitteet ovat keskeisiä valtaosalle korpustutkimusta, mutta niiden käyttö ja tarkka merkitys vaihtelee hieman tutkijoittain. Valotan seuraavaksi ennen kaikkea sitä, miten ymmärrän nämä käsitteet tämän väitöskirjatutkimuksen puitteissa. Esittelyt eivät siis pyri olemaan tyhjentäviä.

n-grammit (ks. esim. Jurafsky & Martin 2009: 83) eli monisanaiset ketjut (engl. *multi-word strings*, Mauranen 2000: 120) klusterit (engl. *cluster*, Kenny 2000) tai leksikaaliset kimput (engl. *lexical bundles*, Biber & Conrad 1999) ovat kielen elementtien ketjuja, joiden pituus on n. Esimerkissä 17 on siis neljä 1-grammia (*kuka*, *potkaisi*, *pallon* ja *pihalle*), kolme 2-grammia (*kuka potkaisi*, *potkaisi pallon* ja *pallon pihalle*), kaksi 3-grammia (*kuka potkaisi pallon* ja *potkaisi pallon pihalle*) ja yksi 4-grammi (*kuka potkaisi pallon pihalle*).

17)	Kuka	potkaisi	pallon	pihalle?
lemma	<kuka>	<potkaista>	<pallo>	<piha>
pos	<pron>	<v>	<n>	<n>
mrp	<sg nom>	<fin ind pret sg3>	<sg gen>	<sg all>
fun	<npsubj>	<pred>	<npobj>	<advl>

Valtaosassa aiempaa korpustutkimusta grammien katsotaan koostuvan sanoista (mm. yllä mainitut tutkimukset). Grammit voivat kuitenkin olla yhtä lailla esimerkiksi aineiston kielelliseen annotaatioon eli lingvistiseen metatietoon perustuvia sanaluokka- (esimerkissä pos), muoto- (esimerkissä mrp) tai funktiomerkitöjä (esimerkissä fun) (ks. esim. Aarts & Granger 1998; Wiersma ym. 2011). Tällä tavalla ajatellen esimerkissä on tässä tutkimuksessa käytetyn LAS2-korpuksen annotaatiota seuraten (ks. LAS2) neljä sanan perusmuotoa eli lemmaa kuvaavaa 1-grammia (<kuka>, <potkaista>, <pallo> ja <piha>), neljä sanan sanaluokkaa kuvaavaa 1-grammia (<pron> eli pronomini, <v> eli verbi ja kaksi <n>-grammia eli substantiivia), neljä sanan morfologista muotoa kuvaavaa 1-grammia (<sg nom> eli yksikön nominatiivi, <fin ind pret sg3> eli finiittimuotoinen indikatiivin preteritin kolmas persoona, <sg gen> eli yksikön genetiivi ja <sg all> eli yksikön allatiivi) ja neljä sanan

syntaktista funktiota kuvaavaa 1-grammia (<npsubj> eli nominaalinen subjekti, <pred> eli predikaatti, <npobj> eli nominaalinen objekti ja <advl> eli adverbiaali). Vastaavasti myös 2-, 3- tai 4-grammeja voisi tarkastella kaikilla näillä annotaation eri tasoilla. Kuten edellä näkyy, morfologiset grammit koostuvat monesti useasta morfeemista. Sanojen morfologiaa tarkasteltaessa olisikin mahdollista tarkastella grammeja myös morfeemeittain siten, että jokainen morfeemi vastaa yhtä morfologisen tason grammia. Katson tässä tutkimuksessa yhden morfologisen grammin kuitenkin koostuvan nimenomaan yhden sanan koko morfologisesta annotaatiosta. Yhtäältä tähän on ohjannut käytetyn korpuksen annotaatio ja sanan rooli annotaation perusyksikkönä (ks. LAS2): annotaatio on kohdistettu sanoittain, eikä morfeemeja ole eroteltu tai niitä vastaavaa annotaatiota kohdistettu tarkemmin morfeemeittain. Morfologisessa annotaatiossa huomioidaan lisäksi myös ns. nollamorfiin (kuten yksikölliset nominit esimerkiksi 17) kaltaisia seikkoja, joilla ei ole osoitettavissa olevaa morfeemia. Toisaalta myös annotaation eri tasojen välinen vertailu ja rakenteiden analogisten piirteiden tarkastelu olisi huomattavasti hankalampaa, mikäli kukin morfologisen annotaation osa käsitettäisiin yhdeksi grammiksi: on helppoa sanoa 2-grammien <sg nom><fin ind pret sg3> ja <npsubj><pred> kuvaavan samaa aineiston sekvenssiä, kun taas yksittäisten morfologisten tunnusten suhdetta annotaation muihin tasoihin on tämän aineiston puitteissa vaikea tarkastella.

Käsite n-grammi ei ota itsessään kantaa elementtien laatuun, vaan keskeistä on niiden järjestys ja peräkkäisyys. Joissakin määritelmässä n-grammit erotetaan kollokaatioista toteamalla, että kollokaatiot ovat kahden sanan pareja ja klusterit useampisanaaisia yhdistelmiä (ks. esim. Mauranen 2000: 120). Yleensä n-grammeja tarkasteltaessa pyritään löytämään samanlaisina tai samankaltaisina toistuvia n-grammeja ja näin löytämään kielessä toistuvia malleja, ja esimerkiksi Dorothy Kennyn mukaan käsite merkitsee jo itsessään toistuvaa myötäesiintymäkokonaisuutta (Kenny 2000: 99). Tämän väitöskirjan puitteissa n-grammit ovat minkä tahansa kielellisten elementtien yhteisesiintymien joukkoja riippumatta niiden yhteisesiintymisen yleisyydestä. Käytännössä tarkasteluun kuitenkin nousee vain sellaisia n-grammeja, jotka esiintyvät yhdessä verraten usein tai joiden esiintymisfrekvenssissä on kulloinkin tarkasteltavien osa-aineistojen välillä selvästi havaittava ero.

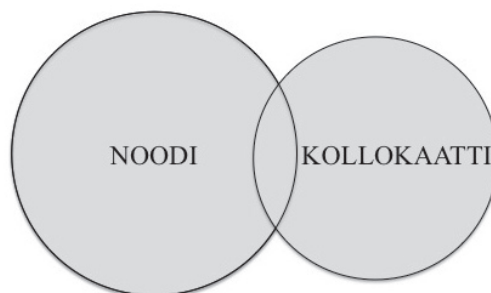
Edellä kuvatuille n-grammien määritelmille on yhteistä se, että grammin jäsenien odotetaan olevan peräkkäin ja samassa järjestyksessä. Kuten Sinclair toteaa, tarkasteltavassa

kielenpiirteessä esiintyvä vaihtelu tai moniosaisuus voi näin jäädä huomioimatta (Sinclair 2001: 353). Grammeja voi kuitenkin määrittellä myös siten, että sanojen ei tarvitse olla peräkkäin, kunhan ne ovat samassa järjestyksessä. Tässä niin sanotussa skipgrammi-lähestymisessä (engl. *skipgram*) sanojen väliin voi grammien esiintymisfrekvenssejä laskettaessa jäädä tutkijan määrittelemä määrä sanoja (Guthrie ym. 2006). Esimerkissä 17 olisi tällä tavoin laskien kolmen 2-grammin sijaan kuusi 2-grammia (*kuka potkaisi, kuka pallon, kuka pihalle, potkaisi pallon, potkaisi pihalle ja pallon pihalle*). David Guthrie ym. (2006) osoittavat, että skipgrammien avulla taustamuuttujiltaan samanlaisten tekstien (kyseisessä tutkimuksessa British National Corpus -korpuksen sanomalehtitekstien) erottaminen taustamuuttujiltaan toisenlaisista teksteistä (kyseisessä tutkimuksessa kiinasta englantiin automaattisesti käännettyistä sanomalehtiteksteistä) parani huomattavasti, kun erotteluun käytettiin tavallisten n-grammien sijaan skipgrammeja (Guthrie ym. 2006: 1224–1225). Vastaavasti sanojen yhteisesiintymistä voidaan tarkastella myös konkgrammein (engl. *concgram*), jolloin sanojen yhteisesiintymisen frekvenssiin ei vaikuta niiden keskinäinen järjestys, vaan ainoana kriteerinä on tutkijan määrittämä sijainnin läheisyys (Cheng ym. 2006). Tällä tavoin esimerkiksi grammit *potkaisi pallon* ja *pallon potkaisi* lasketaan saman n-grammin esiintymiksi. LAS2-korpuksen hakutyökalujen avulla grammifrekvenssejä voidaan laskea joko peräkkäisten grammien tai skipgrammien osalta (ks. LAS2).

Käytän tässä väitöskirjatutkimuksessa käsitettä n-grammi etsiessäni sellaisia yhden, kahden tai kolmen sanan pituisia kielen toistuvia malleja, joiden frekvenssi erottaa S1- ja S2-oppijoiden tekstejä toisistaan (ks. *S2-avainrakenteet*), eri ensikielisiä S2-oppijoita toisistaan (ks. *Siirtovaikutus-avainrakenteet*) tai joiden käyttöfrekvenssi muuttuu selvästi pitkittäistarkastelun aikana (ks. *Pitkittäismuutokset*). Lasken näissä tutkimuksissa aineistosta kaikkien siinä esiintyvien n-grammien frekvenssin ja kohdistan analyysin niihin grammeihin, joissa on suurin frekvenssiero. Näen nämä erottavat toistuvat mallit potentiaalisiksi avainmuodoiksi, joiden avulla voidaan päästä käsiksi konstruktionaalisiin eroihin (ks. tarkemmin 3.2) eli avainrakenteisiin. Tarkastelun lähtökohtana ovat muotorakenteessa havaittavat erot. Tarkastelen siis n-grammien avulla morfologisten n-grammien esiintymisfrekvenssejä. Esiintymisfrekvenssejä laskettaessa olen päätenyt käyttämään skipgrammi-lähestymistä. Tällä tavoin vertailussa voi helpommin nousta esiin myös sellaisia kompositionaalisia ilmiöitä, joissa grammien välillä on keskinäinen vaikutussuhde, vaikka ne eivät esiinny peräkkäin. Guthrien ym. (2006) tulkintoja seuraten on luontevaa olettaa, että aineistojen väliset erot nousevat laajemmin esiin itse aineistoa kasvattamatta, kun grammien

frekvenssien tarkastelussa sovelletaan skipgrammi-lähestymistä. En ole tarkastelussa määritellyt sanojen välistä maksimietäisyyttä, mutta niiden on oltava samassa lauseessa. Näin ollen voitaneen olettaa, että havaittavat erot kuvaavat ennen kaikkea lausetason ilmiöitä. n-grammin kaikkien grammien on oltava samassa järjestyksessä, jotta ne lasketaan saman grammin esiintymiksi. Tästä syystä erot esimerkiksi jonkin tietyn konstruktion sanajärjestyksessä nousevat esiin nimenomaan sanajärjestyksessä havaittavan eronsa ansiosta.

Kollokaation käsite on tullut tutuksi ennen kaikkea John Firthin myötä (Firth 1968: 179–181). Myös kollokaatit ovat toistuvia malleja, sillä ne kuvaavat sanojen yhteisesiintymistä. Tarkastelunäkökulma kuitenkin poikkeaa n-grammeista: nimensä mukaisesti lähtökohtana on jokin sana eli noodi, jonka kanssa usein esiintyviä sanoja voidaan kutsua sen kollokaateiksi. Kollokaatio on siis syntagmaattinen eli kielenyksiköiden peräkkäisyyttä kuvaava esiintymissuhde, jolla kuvataan noodin suhdetta kollokaatteihinsa. Useimmiten kollokaatiota kuvataan tarkastelemalla noodin läheisyydessä, Nadja Nesselhaufin (2004: 12) mukaan yleensä korkeintaan neljän sanan päässä noodista esiintyviä sanoja. Etäisyyden määrittely lienee ainakin osin käytettävissä olevien työkalujen sanelemaa. Kiinnostus kollokaatioita kohtaan on kuitenkin kummunnut leksikon tutkimuksesta ja siitä, miten kielenkäyttäjät valitsevat sanoja yhtä sanaa suurempina joukkoina (ks. esim. Sinclair 1991: 115), mistä syystä sanojen välinen etäisyys on perinteisesti ollut verraten pieni. Sinclairin (mp.) mukaan kollokaatiosuhde on tapa jäsentää alaluvussa 2.1.3 esiteltyä idiomiperiaatetta eli kielenpiirteiden yhteisesiintymisen odotuksenmukaisuutta ja tyypillisyyttä.



Kuvio 1. Noodin ja kollokaatin välinen kollokaatiosuhde.

Kuviossa 1 vasemmanpuoleisen ympyrän pinta-ala kuvaa noodin esiintymäfrekvenssiä, oikeanpuoleisen ympyrän pinta-ala sen mahdollisen kollokaatin esiintymäfrekvenssiä ja ympyröiden päällekkäinen alue siis niitä tapauksia, jolloin noodi ja mahdollinen kollokaatti

esiintyvät yhdessä tarkastelualueella. n-grammeista poiketen kollokaatiosuhteella ei kuvata pelkästään kahden sanan yhteisesiintymistä, vaan myös tämän yhteisesiintymisen suhdetta noodin ja kollokaatin omaan esiintymäfrekvenssiin, eli kuviossa 1 ympyröiden päällekkäisen alueen suhdetta ympyröiden omaan pinta-alaan. Noodin ja sen kollokaatin kollokaatioarvo eli kollokaation vahvuus on sitä suurempi, mitä suuremman osan noodin ja kollokaatin yhteisesiintymät kattavat niiden esiintymisestä yksinään. (Sinclair 1991: 115–119.) Sinclair (1987b: 325–326) huomauttaa lisäksi kollokaation voivan olla luonteeltaan neutraalia, jolloin noodin ja kollokaatin esiintymäfrekvenssit ovat kutakuinkin samankaltaiset, ylöspäin suuntautuvaa, jolloin kollokaatti on selvästi noodia frekventimpi, tai alaspäin suuntautuvaa, jolloin kollokaatti on selvästi noodia harvinaisempi.

Fraseologisten merkitysyksiköiden ja tilastollisesti todennäköisen yhteisesiintymisen erottamisen tapaan (ks. 2.1.3) Nesselhauf jaottelee kollokaatioiden määritelmät kahtia statistisiin ja merkityspenastaisiin lähestymistapoihin: ensinnäkin kollokaatioita ovat sellaiset sanat, joiden esiintyminen vierekkäin tai toistensa läheisyydessä on todennäköisempää kuin vastaavien sanojen tällainen esiintyminen keskimäärin, ja toisaalta kollokaatioita ovat sellaiset sanat, joiden yhtymä on jossain määrin kiinteä ja mahdollisesti myös merkityseron synnyttävä tekijä. (Nesselhauf 2004: 11–12.) Suomen kielen tutkimuksessa esimerkiksi Jantunen soveltaa sanojen kollokaatiosuhteiden tarkastelua tutkiessaan astemääritteiden *oikein*, *kovin* ja *hyvin* kontekstuaalisia käyttöeroja natiivisuomen ja käännösuomen aineistoissa (Jantunen 2004) sekä tarkastellessaan *aivan*-sanana tyypillistä käyttöympäristöä sekä lähisyronymien *liikahtaa* ja *hievahtaa* eroja sanojen semanttisissa preferensseissä (Jantunen 2009b). Kollokaation käsite ei itsessään ota kantaa noodin ja kollokaatin järjestykseen. Usein järjestyksen oletetaan kollokaatioita tarkastelevassa tutkimuksessa implisiittisesti olevan kiinteä, jotta suhdetta voidaan tarkastella kollokaationa, ja esimerkiksi Nesselhauf (2004) tutkii ainoastaan oppijanenglannin verbi–substantiivi-kollokaatioita (esim. *make an attempt*). Vaihtoehtoisesti noodin ja sen kollokaattien suhdetta voidaan tarkastella siten, että sijainti on yksi kollokaatiosuhdetta kuvaava tekijä. Tätä ajatusta seuraten esimerkiksi Jantunen toteaa suomen *aivan*-sanana vahvimpien kollokaattien sijaitsevan nimenomaan noodin oikeanpuoleisessa kontekstissa (Jantunen 2009b: 362–366). (Yhteenvetona kollokaatio-käsitteen erilaisista tulkinnoista, ks. esim. Jantunen 2004: 16–21.)

Käsitän kollokaation myös tässä tutkimuksessa noodin ja sen kanssa usein esiintyvän leksikaalisen aineksen väliseksi suhteeksi. En pidä kollokaation ja seuraavaksi esiteltävän

kolligaation eroa minään perustavanlaatuisena jakona, joskin kollokaation sisäinen variaatio eli tyyppifrekvenssi on sen leksikaalisen luonteen vuoksi vähäisempää kuin kieliopillisten elementtien yhteisesiintymistä kuvaavan kolligaation. Tutkimustenvälisen verrattavuuden ja terminologisen yksiselitteisyyden vuoksi viittaa kollokaatiolla nimenomaan sanastollista yhteisesiintymistä kuvaavaan suhteeseen. Nähdäkseni noodin ja sen kollokaatin välinen järjestys on syytä ottaa huomioon suhdetta tarkasteltaessa, sillä järjestys kuvaa usein myös noodin ja kollokaatin välistä merkityssuhdetta. Suomen kielen kannalta lienee esimerkiksi odotuksenmukaista, että noodin ollessa substantiivi ovat sen vasemmanpuoleiset nominikollokaatit todennäköisimmin noodin määritteitä, kun taas oikeanpuoleiset kollokaatit ovat useimmiten noodin pääsanoja. Esimerkiksi lausekkeissa *lapsen laulu* ja *laulun lapsi* kokonaismerkitys on hyvin erilainen, vaikka sanojen perusmuotoja tarkasteltaessa niiden kollokaatiosuhde näyttäytyisi samanlaisena, mikäli järjestystä ei oteta huomioon.

Kolligaation käsite on kollokaation tavoin tullut tunnetuksi ennen kaikkea Firthin myötä. Siinä missä kollokaatio kuvaa lekseemien välistä syntagmaattista valintapreferenssiä, kolligaatio luonnehtii erilaisten kieliopillisten luokkien tai niiden jäsenten välistä vastaavanlaista valintapreferenssiä. (Firth 1968: 181–183.) Tätä eroa lukuun ottamatta edellä kuvatut, kollokaatiosuhdetta jäsentävät seikat ovatkin nähdäkseni sovellettavissa yhtä lailla myös kolligaatiosuhteiden tarkasteluun. Kolligaatio-termin merkitys on laajentunut tarkoittamaan kieliopillisten luokkien syntagmaattisten suhteiden lisäksi myös jonkin lekseemin ja sen kielellisessä lähikontekstissa esiintyvän kielioppikategorian välistä suhdetta (ks. esim. Stubbs 2001: 449). Elena Tognini-Bonellin mukaan kolligaatiolla tarkoitetaan etenkin korpuslingvistiikassa usein juuri tätä (Tognini-Bonelli 1996: 74), mutta toisaalta esimerkiksi Susan Hunstonin ja Gill Francisin (2000) kieliopin toistuvien mallien tarkastelussa malleja luonnehditaan niin kieliopin kategorioiden välisenä suhteena (esim. verbi-nomini-substantiivi-grammina, mts. 87) kuin lekseemien ja kieliopillisten kategorioiden välisenä suhteena (esim. adjektiivi-*that*-grammina, mts. 93). Lisäksi niin Graeme Kennedy (1991), Michael Hoey (1997: 8) kuin Jantunenkin (2001, 2004) pitävät tarkasteltavan elementin lauseasemaa mahdollisena kolligaatiosuhteena (yhteenvetona kolligaation käsitteen erilaisista tulkinnoista ks. esim. Jantunen 2004: 21–24). Näin esimerkin 17 lauseessa *Kuka potkaisi pallon pihalle?* sanan *Kuka* välittömästi edeltävänä vasemmanpuoleisena kolligaattina on lauseenalku, kun taas välittömästi seuraavana oikeanpuoleisena kolligaattina on sanaluokkatasolla verbi, ja morfologisella tasolla yksikön kolmannen persoonan indikatiivin preteriti ja syntaktisella tasolla predikaatti. Pidän tämän tutkimuksen puitteissa kolligaatioina

kaikkia sellaisia tyypillisiä esiintymisympäristöjä kuvaavia suhteita, joissa ainakin toinen elementti määrittellään sen edustaman kieliopillisen luokan (tässä väitöskirjassa ennen kaikkea sanaluokan, morfologisen muodon tai syntaktisen funktion, ks. LAS2) tai lauseaseman perusteella.

Kollostruktio-käsitteen avulla pyritään yhdistämään konstruktiokieliopin (ks. 2.1.2) teoreettisen viitekehyksen ja korpustutkimuksessa yleisen kielellisen lähikontekstin analyysin vahvuudet. Kollostruktio-termin ovat esitelleet Stefanowitsch ja Gries (2003), ja se perustuu siihen konstruktiokieliopin peruslähtökohtaan, että kieliopin rakenteet eivät poikkea perustavanlaatuisesti sanoista, vaan ne ovat kaikki konstruktioita, joiden skemaattisuus on asteittaista ja yksikkökohtaista. Kuten Gries ja Stefanowitsch huomauttavat, puhtaan lineaarisen kollokaatiosuhteen tarkasteleminen rajoittaa sen sovellettavuutta syntaktisessa kuvauksessa (Gries & Stefanowitsch 2004). Kollostruktioanalyysin tavoitteena onkin tarkastella sitä, miten vahvasti konstruktioiden käyttö assosioituu tiettyihin sanoihin. Käytännössä tämä tarkoittaa sitä, miten usein jokin tietty lekseemi esiintyy jossain tietyssä konstruktiossa ja kuinka suuren osan kyseisen lekseemin ja konstruktion esiintymistä yhteisesiintyminen kattaa. Menetelmällisesti idea vastaa siis kollokaatio- ja kolligaatioanalyysia, mutta noodina ja kollokaattina toimivat joko jokin konstruktio ja siinä esiintyvä vaihtelu tai jokin lekseemi ja ne konstruktiot, joissa kyseinen lekseemi esiintyy. Näin ollen yhteisesiintyminen on kollokaatiosta ja kolligaatiosta poiketen luonteeltaan paradigmaattista eli jonkin konstruktion sisäistä vaihtelua kuvaavaa.

Stefanowitsch ja Gries (2003) osoittavat, että kollostruktiosuhdetta tarkastelemalla pystytään kuvaamaan luontevasti ilman ennakko-oletuksia esimerkiksi ditransitiivirakenteen kaltaisten skemaattisten konstruktioiden tyypillistä merkitystä (mts. 227–230) ja toisaalta esimerkiksi tiettyjen sanojen esiintymistä erilaisissa konstruktiossa ja siihen liittyvää konstruktiokohtaista polysemiaa (mts. 220–222). Ajatus vastaa monilta osin konstruktiokieliopin käyttöpohjaista perustaa ja sitä, että skemaattistenkin konstruktioiden kokonaismerkitys on usein lähellä siinä prototyypisimmin esiintyvien lekseemien merkitystä (ks. Goldberg ym. 2004). Suomen kannalta kollostruktiosuhdetta voi havainnollistaa vaikkapa eksistentiaalikonstruktion [NP_{INE} V_{SG3} NP_{NOM/PART}] avulla. Kuten Iso suomen kielioppi toteaa (VISK §893), verbinä on useimmiten *olla* mutta siinä voi esiintyä myös muu intransitiivinen ja merkitykseltään tyypillisesti olemassa olemista tai olemaan tulemista kuvaava verbi, kuten *tulla* tai *ilmestyä*. Koska *olla*-verbillä on kuitenkin myös paljon

muunlaista käyttöä, voidaan ajatella, että muut olemassa olemista kuvaavat verbit assosioituvat eksistentiaaliseen konstruktion jona *olla*-verbiä vahvemmin.

Aiempi tutkimus tarkastelee kollostruktiosuhteen avulla lähinnä jonkin konstruktion ja siinä esiintyvän leksikon välistä suhdetta (ks. esim. Stefanowitsch & Gries 2003; Gries & Stefanowitsch 2004; Hilpert 2006; Gries & Hilpert 2008; Mukherjee & Gries 2009). Kuten todettua, konstruktiokieliopin kannalta skemaattisuudeltaan erilaiset konstruktiot eivät kuitenkaan ole perustavanlaatuisesti erilaisia. En näekään mitään syytä, miksi tarkastelua ei voisi laajentaa kuvaamaan myös konstruktioiden ja niiden osissa esiintyvän muun kuin leksikaalisen vaihtelun välistä suhdetta. Suomen kielen kaltaisessa, morfologisesti rikkaassa kielessä tuntuu luontevalta laajentaa tarkastelu joissain tapauksissa esimerkiksi jonkin konstruktion morfologiseen variaatioon ja sen kautta havaittaviin myötäesiintymiin. Tällaisia suhteita voisivat olla vaikkapa jonkin konstruktion jossakin osassa havaittavat pääluokan, luvun tai tempuksen konstruktiokohtaiset assosiaatiot. Käsitän siis tämän väitöskirjatutkimuksen puitteissa kollostruktio-termin kattavan minkä tahansa konstruktion ja sen jossakin osassa havaittavan vaihtelun välisen esiintymissuhteen tarkastelun.

2.2. Toisen kielen oppiminen

Käsitteellä toinen kieli on käyttökontekstista riippuen monia eri käyttötapoja. Keskeisimpänä tekijänä lienee se, että toisen kielen oppija tai käyttäjä on oppinut jonkin muun kielen ennen tarkastelunalaista kielen oppimista. Usein hienojakoisempi erottelu on kuitenkin välttämätöntä, ja mahdollisia erontekojä voivat olla esimerkiksi oppimistapa tai -paikka (*toinen kieli* vs. *vieras kieli*, ks. esim. Lomaa–Tuomela 1993: 239) tai opittujen kielten määrä tai järjestys (*toinen kieli* vs. *kolmas kieli* tai *n:s kieli*, ks. esim. Cenoz ym. 2001). Toisaalta monikielisyuden tutkimus on ylipäätään osoittanut, että opittujen kielten määrä ja oppimisjärjestys yhdessä kieltenvälisten vaikutusten kanssa eivät välttämättä aina ole perusteltuja jäsentämistapoja kielitaitoa ja kielenoppimista tarkasteltaessa (ks. esim. Blommaert & Rampton 2011).

Yhtenä tämän väitöskirjan tavoitteista on kuitenkin tutkia sellaisia tyypillisiä ja toistuvia seikkoja, jotka nousevat esiin aineistosta itsestään ja jotka yleisesti yhtäältä erottavat ensikielisten ja ei-ensikielisten kielenkäyttäjien suomea toisistaan (ks. *S2-avainrakenteet*), indikoivat ensikielikohtaisia eroja ei-ensikielisten suomenkäyttäjien välillä (ks. *Siirtovaikutus-avainrakenteet*) tai jotka muuttuvat ajan kuluessa vielä edistyneidenkin suomenoppijoiden kielessä (*Pitkittäismuutokset*). Tällä tavoin tarkempi tutkimus voidaan kohdistaa aineistosta

esiin nouseviin, tarkasteltavia kielimuotoja toisistaan erottaviin kielen ilmiöihin (ks. *Mahdollisuuden ilmaukset*). Kaikki tämän tutkimuksen suomi toisena kielenä -aineiston (S2-aineisto) tekstit ovat sellaisten kielenkäyttäjien kirjoittamia, joiden ensikieli on jokin muu kuin suomi ja jotka ovat oppineet suomea vasta lapsuusiän jälkeen. Vastaavasti suomi ensikielenä -aineiston (S1-aineisto) tekstit puolestaan ovat sellaisten kielenkäyttäjien kirjoittamia, joiden ensikieli on suomi ja jotka ovat saaneet koulusivistyksensä suomeksi. Puhuessani ei-ensikielisistä tai suomea toisena kielenä käyttävistä kielenkäyttäjistä tai suomenoppijoista viittaankin yksinkertaistaen siihen, että suomi ei ole ensimmäinen opittu kieli, en oppijoiden kielitaitotasoon tai ensikielen jälkeen opittujen kielten määrään ja oppimisjärjestykseen. Pyrin tutkimuksessani tarkastelemaan sitä, onko ei-ensikielisten suomenkäyttäjien ja suomea ensikielenään oppineiden kielenkäyttäjien suomessa havaittavissa sellaisia toistuvia eroja, jotka korreloivat edellä kuvatun ensikielisyys-taustamuuttujan kanssa, sekä sitä, millaisia nämä mahdolliset erot ovat luonteeltaan. Havaittavat erot voivat paljastaa sellaisia kielellisiä ilmiöitä, jotka ovat ominaisia nimenomaan suomelle, tai sellaisia ilmiöitä, joissa oppijoiden muu kielitaito vaikuttaa myös heidän käyttämäänsä suomeen. En kuitenkaan oletta näiden kielenkäyttäjien suomen välttämättä eroavan lähtökohtaisesti perustavanlaatuisesti ensikielisestä suomesta. Kyse on käyttöpohjaisen kielikäsitelmän hengessä nähdäkseen ennemminkin eroista kaiken aiemmin kohdatun kielen määrässä ja laadussa. Yhtenä muuttujana tässä aiemmin kohdatun kielen kokonaisuudessa on kuitenkin luonnollisesti myös kielenkäyttäjien aiempi kielitaito ja heidän aiemmat kokemuksensa tämän aiemman kielitaidon mukaisista tyypillisistä kielenkäytön mekanismeista ja näiden mekanismien toiminnasta erilaisissa kielenkäytön konteksteissa. Karkeasti yleistäen erottavia tekijöitä voivat siis olla aiemmat kieliympäristöt, ja näiden aiempien kieliympäristöjen sisältämistä rekisteri- tai tekstilajikohtaisista tavoista ja muusta aiemmasta kielitaidosta kumpuavat erot kielellisissä kokemuksissa ja kielenkäyttötavoissa.

Esittelen tässä alaluvussa 2.2 sitä, miten näen toisen kielen oppimisen ja suomen toisena kielenä niveltävän käyttöpohjaiseen kielikäsitelmään ylipäätään (ks. luku 2.1). Yleisten teoreettisten suuntaviivojen lisäksi keskityn erityisesti siihen, miten konstruktiolähtöistä näkemystä kielestä voidaan soveltaa toisen kielen oppimiseen ja erityisesti suomeen toisena kielenä. Kaikki S2-aineiston informantit ovat edistyneitä suomenoppijoita, mistä syystä heidän opiskelutaustansa ovat niin ikään hyvin moninaiset. Kaikki informantit ovat kuitenkin opiskelleet suomea sekä Suomessa että suomen ulkopuolella, mistä syystä en tämän tutkimuksen puitteissa näe mielekkäänä erottaa suomea toisena ja vieraana kielenä tai

luokkahuoneympäristössä hankittua kieitaitoa ja kohdekielisessä ympäristössä hankittua kielitaitoa toisistaan.

2.2.1. Ensimmäisen ja toisen kielen välinen suhde

Yksi keskeisistä toisen kielen tutkimuksen kysymyksistä on se, eroavatko ensimmäisen ja toisen kielen oppimisen prosessit toisistaan jollain lailla perustavanlaatuisesti vai eivät. En varsinaisesti käsittele asiaa tämän väitöskirjan osatutkimuksissa, mutta tutkimukseni käsittelee kuitenkin myös konstruktioiden luonnetta kielen yksiköinä. Pohdin lisäksi niitä kontekstuaalisia syitä, jotka vaikuttavat konstruktioiden oppimiseen. Näiden syiden vuoksi katson tarpeelliseksi pohtia sitä, ovatko kielen yksiköt ja kielioppi perusluonteeltaan samanlaisia toisella kielellä toimivilla ja ensikielisillä kielenkäyttäjillä. Näkemystäni seuraa käyttöpohjaista kielikäsitystä, ja ajattelen, että oppimismekanismit eivät välttämättä itsessään eroa radikaalisti toisistaan mutta että tilanteiden erilaisuuden vuoksi myös prosessit saattavat poiketa toisistaan. Valotan siis tässä aluvuossa sitä, millaisena näen ensimmäisen ja toisen kielen oppimisen välisen suhteen.

Nick Ellis (2003: 72) nostaa esiin kolme keskeistä eroa ensikielen ja toisen kielen oppimisen välillä. 1) Oppijan konseptuaalinen kehitys on hyvin erilaista, sillä ensikielen oppiminen tapahtuu tyypillisesti yhtä aikaa kognitiivisten taitojen ja maailmantiedon karttumisen kanssa. 2) Ensikieltä opittaessa tyypillinen kielellinen syötös on luonnollista ja se toistaa samoja lapsen fyysisen kehityksen luomia puitteita, kun taas toisen kielen kohdalla esimerkiksi luokkahuone voi muuttaa, häiritä tai jopa rikkoa luonnollisen altistumisen, funktionaalisuuden ja sosiaalisen kanssakäymisen malleja. 3) Toisen kielen oppiminen perustuu aina oppijan tietoon hänen ensikielestään, ja siinä missä ensikielessä leksikaaliset mallit voivat ohjata esimerkiksi morfosyntaktisten skeemojen kehkeytymistä, toista kieltä opittaessa oppijalla on jo tällaisia kategorioita. Nämä ovatkin tekijöitä, jotka on otettava huomioon taustamuuttujina aina, kun tarkastellaan ensikielen ja toisen kielen oppimisen välisiä eroja.

Toisaalta itse kehitys- ja oppimisprosessit eivät välttämättä ole erityisen monimutkaisia, nämä prosessit vain ovat alttiita monimutkaiselle ja kieliaineistoltaan rikkaalle ympäristölle (Ellis, N. 1998: 644). Ajatus vastaa monilta osin luonnontieteissä ja yhteiskuntatieteissä sovellettua kompleksisuusteoriaa (engl. *complexity theory*). Teoria pyrkii kuvaamaan sitä, miten jonkin järjestelmän (kuten ihmiskeho tai tietty ihmisryhmä) osat yhdessä synnyttävät

kyseessä olevan järjestelmän tyypillisen toiminnan ja miten tällainen järjestelmä samanaikaisesti toimii vuorovaikutuksessa ympäristönsä kanssa. Yksi tunnetuimmista kompleksisuusteorian esimerkeistä on ns. perhosvaikutus (engl. *butterfly effect*), jonka mukaan perhosen yksittäinen siivenisku voi saada aikaan myrskyn maapallon toisella puolella, kun kaikki toisiinsa vaikuttavien seikkojen rooli näiden tapahtumien välissä otetaan huomioon. Kompleksisuusteorian soveltaminen kielenoppimisen tarkasteluun korostaa holistista näkökulmaa, jonka mukaan oppiminen on luonteeltaan epälineaarinen ja vaihtelulle altis prosessi. Kompleksisuusteorian mukaan yksilön kielellinen järjestelmä kehkeytyy jatkuvassa vuorovaikutuksessa hänen ympäristönsä kanssa siten, että kieli ja kielitaito ovat peruluonteeltaan muuttuvia ja mukautumaan pyrkiviä, toisiinsa kietoutuneita järjestelmiä, jotka saavat muotonsa sen perusteella, miten kieltä käytetään. Ihmisen yleiset imitoimisen, kategorioinnin ja rakenteellisen samankaltaisuuden havaitsemisen kaltaiset kognitiiviset kyvyt ja kommunikatiiviset tarpeet mahdollistavat kielen oppimisen, ja näiden kykyjen ja tarpeiden ohjaamana kieltä opitaan perhosvaikutuksen tapaan. Näin ajatellen kielen oppimisessa ei siis ole päätepistettä vaan se on jatkuva muutosprosessi, joka sisältää muutostenvälisiä suvantovaiheita. (Kompleksisuusteoriasta ja soveltamisesta kieleen, ks. esim. Larsen-Freeman 1997; Larsen-Freeman & Cameron 2008, joskin jo Maisa Martin (1995) on soveltanut implisiittisesti vastaavaa ajattelua S2-tutkimuksessa.)

Kompleksisuusteorian kannalta ensimmäisen ja toisen kielen oppimista erottavat tekijät vastaavat hyvin pitkälti edellä kuvattuja, Ellisin esittämiä peruseroja: järjestelmä itsessään on erilainen, sillä oppija osaa jo vähintään yhtä kieltä ja hän on kognitiivisesti ensikielen oppijaa harjaantuneempi. Toisaalta myös ympäristö, jonka kanssa oppija on vuorovaikutuksessa, poikkeaa toisen kielen oppijalla ensimmäistä kieltään oppivan ympäristöstä – yhtäältä kommunikaation tarpeet ovat toisenlaisia ja toisaalta toisen kielen oppijan kieliympäristöt poikkevat usein monin tavoin ensikielen oppijan kieliympäristöistä. Kuten Nick Ellisiin viitaten edellä totean, oppimista ohjaavien ja oppimiseen johtavien mekanismien ei kuitenkaan tarvitse olla perustavanlaatuisesti toisistaan poikkeavia (ks. myös Ellis & Larsen-Freeman 2006; Larsen-Freeman & Cameron 2008: 133). Käsittelen oppijakohtaisia ja tilannekohtaisia tekijöitä tarkemmin luvuissa 2.2.3 ja 2.2.4 (ks. myös *Siirtovaikutus-avainrakenteet*, *Pitkittäismuutokset* ja *Mahdollisuuden ilmaukset*), mutta kompleksisuusteorian kannalta keskeistä on Diane Larsen-Freemanin ja Lynne Cameronin mukaan ennen kaikkea ensimmäisen ja toisen kielen välinen typologinen suhde ja oppijoiden näkemys tästä suhteesta (engl. *typological proximity and learners' perception of the proximity*),

kielellisten yksiköiden väliset frekvenssierot ja niiden välinen kilpailu ensimmäisessä ja toisessa kielessä (engl. *competition in frequencies*), altistumisen määrä (engl. *amounts of exposure*), oppijan taidot (engl. *proficiency*) ensimmäisessä ja toisessa kielessä, oppijan orientoituminen ja tavoitteet kielen oppimisen suhteen (engl. *orientation and goals*) sekä kulloisenkin kielellisen tehtävän (engl. *task*) kognitiiviset ja sosiaaliset vaatimukset. Kompleksisuusteorian kannalta on kuitenkin olennaista, että nämä tekijät voivat vaikuttaa myös toisiinsa monin eri tavoin – vaikkapa mitätöiden toistensa vaikutuksen tai moninkertaistaen toistensa vaikutusta. (Larsen-Freeman & Cameron 2008: 133.)

”From a complexity theory point of view, not only do we get a more variegated portrayal of [second] language-using patterns, we also get a different, more emic, or learner-centered, account of their development. Learning is not the taking in of linguistic forms by users, but the constant adaptation of their linguistic resources in the service of meaning-making in response to the affordances that emerge in the communicative situation, which is, in turn, affected by learners’ adaptability.”

(Larsen-Freeman & Cameron 2008: 135; lisäys hakasulkeissa kirjoittajan.)

Kuten edellä oleva lainaus kuvaa, kompleksisuusteorian valossa toisella kielellä toimiminen on jatkuvassa muutoksessa olevan järjestelmän sovittamista kulloiseenkin tilanteeseen. Jokainen vuorovaikutustilanne, jokainen tuotettu ilmaus ja kaikki kuultu syötös vaikuttaa kielenkäyttäjän kielijärjestelmään, Larsen-Freemanin ja Cameronin kuvauksessa kielenkäyttäjän kielellisiin resursseihin. Keskeistä on käyttöpohjaisen kieliopin tapaan se, että kompetenssia ja performanssia ei eroteta toisistaan, vaan ne muodostavat yhdessä toisiinsa kietoutuneina kielijärjestelmän. Ajatus vastaa lopulta monilta osin Larry Selinkerin (1972) alun perin esittelemän välikieli-teorian perusargumenttia siitä, että toisen kielen oppiminen on perusluonteeltaan dynaaminen prosessi, jolle ei voida osoittaa päätepistettä. Vastaavasti myöskään kohdekieli ei ole staattinen sääntöjen kokoelma vaan alati muutoksessa ja ajallisessa liikkeessä oleva järjestelmä, mikä korostaa edelleen sekä kielijärjestelmän että sen oppimisen epälineaarista luonnetta (Larsen-Freeman 1997: 147–150). Kielen dynaaminen eli aikajatkumolla muuttuva luonne onkin nähtävissä samanaikaisesti useilla aikatasoilla, kuten kieliopillistumisen ilmiönä (ks. esim. Traugott & Trousdale 2013), kielenkäyttäjän iän ja altistusajan vaikutuksina oppimiseen (ks. esim. Larsen-Freeman 2006) tai edeltävän kielenkäytön synnyttämänä virittymisenä (engl. *priming*) (ks. esim. Ellis, N. 2002), mistä syystä kielen oppimisen kattavan tarkastelun on otettava huomioon yksittäisten kielenpiirteiden tai konstruktioimaisten muoto–merkitys-parien tarkastelun lisäksi niin niiden lingvistinen lähiympäristö kuin laajempi käyttök kontekstikin.

2.2.2. Konstruktiot toisen kielen opittavina yksiköinä

Larsen-Freemanin ja Cameronin kuvaama, kompleksisuusteoriaa mukaileva malli kielenoppimisesta vastaa käyttöpohjaisen mallin perusolettamuksia: oppimisen ajatellaan noudattavan yleisiä kognitiivisia periaatteita, jotka mahdollistavat niin kielen kuin muiden kognitiivisten kykyjen oppimisen (ks. 2.1.1). Ajatus on siis pääpiirteissään sama kuin aiemmin kuvattujen konstruktioiden kohdalla (ks. 2.1.2), sillä oppimismekanismien ajatellaan nojaavan samankaltaisina toistuvien kielen jaksojen ja niiden samanlaisina toistuvien merkitysten havaitsemiseen ja kykyyn tunnistaa rakenteellista samankaltaisuutta ilmausten välillä (Larsen-Freeman & Cameron 2008: 125). Esimerkiksi Nick Ellis (2002) toteaa, että ihmiset pystyvät menestyksekkäästi tunnistamaan kielen frekvenssiin perustuvia ominaisuuksia ja tekemään sen pohjalta analogisia yleistyksiä havaintojen esiintymä- ja tyyppifrekvenssin välisten suhteiden avulla (ks. myös Bybee & Thompson 1997). Larsen-Freeman (2002) toteaa lisäksi frekvenssien havaitsemisen kohdistuvan nimenomaan muoto-merkitys-parien tunnistamiseen, minkä lisäksi oppimiseen vaikuttavat jatkuvasti myös oppija- ja tilannekohtaiset tekijät (ks. tarkemmin 2.2.4 ja 2.2.5).

Kompleksisuusteoria korostaa Goldbergin (1995; 2006, ks. 2.1.2) konstruktiokielioppia enemmän eri konstruktioiden välisiä yhteyksiä todellisessa kielenkäytössä sekä sitä, että todellisen kielenkäytön sosiaalinen ympäristö ja kielenkäyttö sosiaalisena toimintana vaikuttaa niin ikään oppimiseen. Käsillä olevan kaltaisen, luonteeltaan korpusvetoisen tutkimuksen kannalta onkin keskeistä tarkastella yksittäisten konstruktioiden lisäksi niiden tyyppillistä kotekstuaalista käyttöympäristöä eli sitä, minkälaisen muiden konstruktioiden kanssa, osana tai täydennysjakauman vaihtoehtona kulloinkin tarkasteltava konstruktio voi esiintyä tai tyyppillisesti esiintyy (ks. *S2-avainrakenteet*, *Siirtovaikutus-avainrakenteet*, *Pitkittäismuutokset*, ja *Mahdollisuuden ilmaukset*). Lisäksi tässä tutkimuksessa tarkastelun kohteena on verraten spesifissä kielenkäyttötilanteessa tuotettu tekstilaji – akateemiset tenttivastaukset – mikä on syytä ottaa huomioon pohdittaessa aineistosta esiin nousevien konstruktioiden tyyppillistä käyttöä ja S1- ja S2-aineiston välillä havaittavia eroja. Vaikka käsittelen tutkimuksessani monessa kohdin yksittäisiä konstruktioita, käsitän niiden olevan ensisijaisesti osa kielenkäytön laajaa kokonaisuutta, minkä pyrin myös ottamaan huomioon tulkinnoissani.

Tämän tutkimuksen keskiössä ovat edistyneiden S2-oppijoiden kielen rakenteelliset erityispiirteet. Niinpä yksi keskeisiä kysymyksiä toisen kielen oppimisen erityispiirteiden

kannalta on se, onko toisen kielen rakenteiden oppiminen luonteeltaan samanlaista kuin ensikielen rakenteiden kohdalla (ks. edellä). Nick Ellis (2003: 68–72) puoltaa edellä mainituista erottavista tekijöistä huolimatta ajatusta prosessien samankaltaisuudesta ja ehdottaa, että ensikielen tapaan myös toisen kielen rakenteiden oppiminen noudattaa kehityssekvenssiä, jossa leksikaalisista joukoista (engl. *formulae and idioms*) kehittyvä vähitellen rajattuja toistuvia malleja (engl. *limited scope patterns*) ja näistä edelleen produktiivisia konstruktioita (engl. *constructions*). Ajatusmalli tuntuu perustellulta silloin, kun oppijan syötös on kyllin runsasta ja vastaa edes jossain määrin luonnollista kielenkäyttöä. Luonnollisessa prosessissa oppiminen etenee leksikaalisista joukoista kohti rajattuja toistuvia malleja sitä mukaa, kun kohdattujen, keskenään samankaltaisesti käyttäytyvien kollokaatio- ja kolligaatiosuhteiden määrä on kyllin suuri analogian tunnistamiseen ja yleistyksen muodostumiseen.

Eksplisiittisen ja implisiittisen tiedon suhde ja niiden välinen vaikutus saattaa kuitenkin olla tekijä, joka voi vaikuttaa kehityssekvensseihin ainakin joiltain osin (ks. yhteenvetona esim. Ellis, R. 2008: 449–454). Voidaan esimerkiksi ajatella, että eksplisiittinen tieto ohjaa analogian tunnistamista ja skemaattisen yleistyksen muodostumista. Tämä vastaa Jan Hulstijnin ajatusta siitä, että eksplisiittinen oppiminen ja tieto ei itsessään muutu implisiittiseksi, mutta eksplisiittinen tiedon prosessoiminen saattaa kuitenkin synnyttää implisiittistä oppimista (Hulstijn 2002: 205–209). Myös Nick Ellis näyttää lopulta päätyvän tähän näkökulmaan, sillä hänenkin mukaansa eksplisiittinen oppiminen saattaa nopeuttaa implisiittistä prosessointia konstruktioiden tunnistamisen nopeutumisella ja sen vaikutuksella kielen havaittuihin frekvensseihin (Ellis, N. 2005: 306–308). Tällainen vaihtoehtoinen kehityskulku onkin syytä ottaa huomioon, kun tarkastellaan edistyneen oppijansuomen ja ensikielisen suomen rakenteissa havaittavia eroja ja niiden luonnetta. Edistyneet kielenoppijat ovat usein pitkään jatkuneiden kieliopintojen vuoksi erittäin harjaantuneita opittavan kielen observoijia hyvän toiminnallisen kielitaidon lisäksi, mikä saattaa entisestään vahvistaa eksplisiittisen oppimisen merkitystä. Tämän tutkimuksen informantit ovat lisäksi tehneet lingvistisiä opintoja, minkä voi ajatella entisestään kasvattavan havainnoinnin ja eksplisiittisen oppimisen vaikutusta.

Kuten alaluvussa 2.1.2 todetaan, konstruktionäkökulman avulla voidaan siis jäsentää kielen rakennetta käyttöpohjaisesti, ja konstruktiomaisilla yksiköillä voidaan kuvata myös rakenteiden oppimista (ks. esim. Goldberg ym. 2004) ja niiden mentaalisia representaatioita

(ks. esim. Tomasello 1998). Todelliseen kielenkäyttöön perustuva konstruktio-lähestyminen on osoitettu perustelluksi myös ensikielen jälkeen opittujen kielten oppimisen tarkastelussa. Esimerkiksi Stefan Gries ja Stefanie Wulff (2005) osoittavat saksaa äidinkielenään puhuvia englanninoppijoita käsittelevässä tutkimuksessaan, että opittavan kielen konstruktio- ja rakenteelliset roolit ovat keskeisessä roolissa kielen rakenteen jäsentämisessä myös toista kieltä opittaessa. Vertaamalla kielenoppijoiden tekemän lauseentäydennystehtävän tuloksia englannin- ja saksankielisiin korpuksiin he toteavat, että argumenttirakennekonstruktioiden verbikohtainen virittyminen on samanlaista toisen kielen oppijoiden ja ensikielisten kielenkäyttäjien välillä ja että oppijat eivät sovelle ensikielensä eli saksan konstruktioiden tyypillisiä virittyneisyyksiä opittavan kielen eli englannin konstruktioihin (mts. 189–191). Tutkimuksen toinen osa seuraa aiempaa ensikieleltään englanninkielisille tehdyn tutkimuksen mallia siitä, miten informantit järjestävät joukon sanastoitaan ja argumenttirakenteeltaan vaihtelevia lauseita ryhmiin niiden merkityksen perusteella (ks. Bencini & Goldberg 2000). Tulokset osoittavat, että toisen kielen oppijat nojaavat ensikielisten tapaan konstruktion kokonaismerkitykseen ja että konstruktiomerkitseksen rooli lauseen kokonaismerkitystä tulkittaessa saattaa olla toisen kielen oppijoilla jopa ensikielisiä kielenkäyttäjiä suurempi (Gries & Wulff 2005: 192–194).

Gries ja Wulff (2009) osoittavat konstruktioiden jäsentävän toisen kielen toimintaa myös muissa kuin argumenttirakenteen perusteella määritellyissä konstruktioissa. He tutkivat tiettyjen konstruktioiden käyttöä ja oppijoiden virittymistä käytettyihin konstruktioihin sekä konstruktioiden kollostruktionaalista luonnetta eli leksikaalisen vaihtelun suhdetta konstruktion kokonaismerkitykseen vertaamalla ensikielisten kielenkäyttäjien korpuksista kerätyjä tuloksia oppijoiden tekemiin lauseentäydennys- ja kieliopillisuustesteihin. Tulokset osoittavat tarkasteltujen konstruktioiden virittyvän myös oppijankielellä aiemman käytön perusteella, minkä lisäksi tietyt lekseemit sitoutuvat toisia vahvemmin tiettyyn konstruktion. Adele Goldberg ja Devin Casenheiser (2008) päätyvät samankaltaiseen johtopäätökseen tutkiessaan sitä, miten aiemmin tuntemattoman kielen konstruktiomerkitse opitaan aineistolähtöisesti. Goldbergin ja Casenheiserin (mts. 210–211) mukaan vieraan kielen jonkin argumenttirakenteen konstruktiomerkitse voidaan oppia ja se kytetään yleistämään uusiin konteksteihin verraten pienen syötöksen perusteella, ja yksittäisen konstruktion merkityksen oppimista vahvistaa sen kokonaismerkitykselle tyypillisen verbin runsas esiintyminen konstruktiossa.

S2-tutkimuksen parissa vastaavan, kielen käyttöpohjaista oppimista puoltavan argumentin on esittänyt esimerkiksi Maisa Martin. Väitöskirjassaan Martin tarkastelee suomen kielen nominitaivutusta ja osoittaa, että kontekstittomassa testitilanteessa todellisen kielen sanastoon kuuluvat sanat taivutetaan useammin kohdekielen mukaisesti kuin keksityt, muotorakenteeltaan samanlaiset sanat. S1- ja S2-kielenkäyttäjien kielijärjestelmien keskinäistä samankaltaisuutta puoltaa se, että sama tendenssi toistuu Martinin tutkimuksessa sekä S2- että S1-aineistossa, vaikka kohdekielen mukaiset muodot ovatkin niin todellisten kuin keksittyjen sanojenkin osalta yleisempiä S1-aineistossa. (Martin, M. 1995: 88–130.) Havainto puoltaa näin ollen osaltaan konstruktioiden – tässä tapauksessa nominien taivutusmallien – käyttölähtöisyyttä ja abstraktion käyttöpohjaisuutta sekä ensikielisten että ei-ensikielisten kielijärjestelmässä.

Eksplisiittisesti konstruktiokieliopin käsitteiden varaan nojaavaa S2-tutkimusta on harjoitettu viime vuosina etenkin Jyväskylän yliopistossa toimineiden CEFLING- ja TOPLING-tutkimushankkeiden parissa (ks. CEFLING; TOPLING). Hankkeiden keskeisenä tavoitteena on ollut kartoittaa toisen kielen oppimisen tyypillisiä etenemisreittejä eri kielitaitotasoja vertailemalla. Hankkeiden piirissä tehty, suomen kielen oppimista käsittelevä tutkimus perustuu monissa kohdin konstruktio- ja kielikäsitykseen, ja niin tutkimuskysymykset kuin tarkasteltavien kielenyksiköiden määrittelykin nojaavat ennen kaikkea muodon ja merkityksen välisen vuoropuhelun, korrelaation ja niissä esiintyvän variaation tarkasteluun. Näistä lähtökohdista esimerkiksi Marja Seilonen (2013) tutkii väitöskirjassaan epäsuoria henkilöön viittaamisen keinoja, käyttöä ja mekanismeja oppijansuomessa, Mikko Kajander (2013) tarkastelee eksistentiaalikonstruktion käyttöä, variaatiota ja siihen vaikuttavia tekijöitä toisen kielen oppimisen eri vaiheissa, Nina Reiman (2011, 2014) analysoi transitiivikonstruktiota toisen kielen oppimisen eri vaiheissa ja Sanna Mustonen (2015) lokatiivisuuden ilmauksia erilaisten oppijoiden kielenkäytössä ja oppimisen eri vaiheissa.³

Edellä kuvatut tutkimukset ja havainnot tukevat oletusta kielitaidon käyttöpohjaisesta oppimisesta niin morfologisen produktiivisuuden aineistolähtöisen oppimisen osalta kuin toisaalta muoto-merkitys-parien kohdatusta kielestä abstrahoituvien skemaattisten yleistystenkin kannalta. Tulokset kuvaavat karkeistaen ajatellen sitä, miten kieltä opitaan analogisesti havaitsemalla keskenään rakenteellisesti samankaltaisten kielijaksojen

³ Hankkeiden tarkempiin julkaisuluetteloihin ja mm. niiden piirissä tehtyihin pro gradu -tutkielmiin voi tutustua hankkeiden verkkosivuilla (CEFLING; TOPLING).

toistumista ja niiden käyttäytymistä ja merkityksen syntymistä osana kielenkäyttöä (ks. myös. 2.1.1 ja 2.2.1). Konstruktioiden hahmottumisessa käytön perusteella ei näytä olevan perustavanlaatuisia eroja ensikielisten ja toisella kielellä toimivien kielenkäyttäjien välillä, mikä puolustaa käyttöpohjaisen kielikäsitteiden ajatusta syötökseen perustuvasta oppimisprosessista myös toista kieltä opittaessa.

Toisen kielen oppimisessa on siis runsaasti yhtäläisyyksiä ensikielen oppimiseen. Näin ollen onkin tietyt rajoitukset huomioon ottaen perusteltua ajatella, että kielen kehitys ja oppimisprosessit ovat puitteiltaan samanlaisia kummassakin tapauksessa. Käyttöpohjaisessa viitekehityksessä on siis syytä uskoa, että ensikielen tapaan myös toisen kielen rakenteiden prosessointi ja oppiminen perustuvat kykyyn tiedostamatta havainnoida kielellisessä syötöksessä esiintyviä frekvenssitekijöitä ja keskenään jollain lailla samankaltaisten ilmausten toistumista, merkityksen ja rakenteen tasolla ilmeneviä läheisyysvaikutuksia sekä muita kontekstuaalisia vaikutuksia (ks. esim. MacWhinney 2001: 457–459; Ellis, N. 2002). Tämä liittyy vahvasti aiemmin (2.1.3) esiteltyyn idiomiperiaatteeseen eli karkeistaen ajatukseen siitä, että idiomaattinen kielenkäyttö perustuu kielen havaittuihin tyyppisyyksiin niin, että tietyissä konstruktioiden tai tietyissä kielellisissä lähikonteksteissa jotkut valinnat ovat toisia todennäköisempiä, vaikka nämä kaikki olisivatkin mahdollisia. Toisen kielen oppimisen tutkimuksessa idiomiperiaatteen on esitetty vaikuttavan oppimiseen kolmella vaihtoehdoisella tavalla: 1) niin, että toisen kielen oppijat toimivat ensikielisiä enemmän vapaan valinnan periaatteen varassa (Granger 1998); 2) niin, että oppijat saattavat soveltaa ensikielensä tai muista osaamistaan kielistä tuttuja idiomaattisia ratkaisua myös muihin kieliin (Paquot 2008); 3) niin, että idiomiperiaate sinänsä toimii samalla tavalla niin ensikielisessä kielenkäytössä kuin oppijankielissäkin, mutta erilaisen kohdekielisen syötöksen vuoksi kielelliset ratkaisut ovat erilaisia nimenomaan idiomiperiaatteen vuoksi (Vetchinnikova 2014: 215–216). Nähdäkseni ainakin toinen ja kolmas näkökulma vastaavat edellä esitetyissä, konstruktiokielioppiin nojautuvissa tutkimuksissa saatuja tuloksia. Konstruktioiden oppiminen nojautuu myös toisen kielen oppijoilla konstruktion tyyppisimpään syötöksessä havaittuun käyttöön. Koska analoginen oppiminen tapahtuu käyttöpohjaisesti, vaikuttavat mahdolliset erot saadussa syötöksessä myös siihen, miten konstruktiot opitaan, sekä siihen, millainen on kunkin konstruktion tyyppinen merkitys ja käyttöyhteys. Toisaalta kaikki aiempi syötös vaikuttaa yksilön kielijärjestelmään, mistä syystä myös oppijoiden aiemmalla, niin ikään syötökseen perustuvalla kielitaidolla on osansa.

2.2.3. Oppijakohtaiset tekijät

Kuten edellä jo mainittiin, on yksittäisiä kieleen ja sen oppimiseen vaikuttavia tekijöitä niin paljon, että kattava huomioonottaminen on käytännössä mahdotonta. Tämän väitöskirjan ensisijaisena tavoitteena on tavoittaa ja tarkastella sellaisia piirteitä, jotka ovat yleisiä ja tyypillisiä edistyneen oppijansuomen erityispiirteitä – siis ilmiöitä, jotka kuvaavat suomea toisena kielenä useimpien kielenoppijoiden osalta. Tutkimuksen osatavoitteena on kuitenkin metodologinen tarkastelu sen osalta, miten tällaisia piirteitä voidaan lähestyä erilaisista oppijankielen tutkimukselle tyypillisistä näkökulmista. On esimerkiksi kiintoisaa nähdä, miten jonkin taustamuuttujan – kuten ensikielen – osalta erilaiset oppijaryhmät eroavat toisistaan ja miten tällaisia eroja voidaan tavoittaa aineistosta ilman etukäteen muotoiltua tarkkaa hypoteesia mahdollisten erojen laadusta (ks. *Siirtovaikutus-avainrakenteet*). Esittelen seuraavaksi keskeisiltä osin kieltenvälisiä vaikutussuhteita käsittelevää tutkimusta ja asemoin oman tutkimukseni sen osaksi, minkä lisäksi tarkastelen lyhyesti muita oppijakohtaisia tekijöitä käsittelevää tutkimusta ja sen suhdetta nyt käsillä olevaan väitöskirjatutkimukseeni.

Omaksi luokakseen yksilöllisten oppijakohtaisten taustamuuttujien ja kaikkia oppijoita yhdistävien taustamuuttujien välillä voidaan erottaa ns. kieltenväliset vaikutussuhteet (engl. *crosslinguistic influences* (Sharwood Smith–Kellerman 1986: 1) tai *transfer* (Ellis, R. 2008: 351)). Kieltenvälisiä vaikutuksia ovat tyypillisesti oppijan ensikielen vaikutukset, mutta käsitejoukko kattaa kaikkien muidenkin kielten vaikutukset tarkasteltavana olevaan välikieleen (ks. esim. Cenoz–Hufeisen–Jessner 2001). Esimerkiksi Nick Ellis muistuttaa ensikielen olevan useimmiten suurin syy siihen, että toisen kielen oppija ei ota vastaan kohtaamaansa syötöstä. Tämä johtuu Ellisin mukaan ns. negatiivisesta siirtovaikutuksesta (engl. *interference*), varjostamisesta (engl. *overshadowing*), estämisestä (engl. *blocking*) ja perkeptuaalisesta oppimisesta (engl. *perceptual learning*). Negatiivisella siirtovaikutuksella tarkoitetaan sitä, että oppija siirtää kohdekielen vastaisesti jonkin toisen osaamansa kielen piirteiden tarkasteltavaan kieleen. Varjostamisessa oppijan on tyypillisesti tehtävä kielellinen ratkaisu kahden eri syötöksen herättämän assosiaation väliltä, missä tapauksessa oppija valitsee sen vaihtoehdon, joka on lähempänä hänen ensikieltään. Estäminen puolestaan on automaattisesti opittua ohittamista eli esimerkiksi sellaisen variaation huomioimatta jättämistä, joka ei ole tuttua muista oppijan hallitsemista kielistä. Perkeptuaalisella oppimisella puolestaan tarkoitetaan sitä, että oppija kiinnittää vähemmän huomiota sellaisiin kielenpiirteisiin, jotka ovat redundantteja jossakin hänen aiemmin osaamassaan kielessä

mutta merkityksellisiä tarkasteltavassa kielessä. (Ellis, N. 2006: 176–179.) On kuitenkin hyvä muistaa, että kieltenvälisen vaikutuksen ei välttämättä tarvitse olla oppimista haittaavaa. Tällä perusteella kieltenväliset vaikutukset voidaan Rod Ellisiä (2008: 354–359) mukailleen luokitella edelleen neljään alajoukkoon: 1) virheisiin eli negatiivisiin siirtovaikutuksiin, 2) helpotuksiin eli positiivisiin siirtovaikutuksiin, 3) välttämiseen ja 4) liikakäyttöön. Kieltenväliset siirtovaikutukset voivat olla esimerkiksi suomenkielisillä vironoppijoilla ja vironkielisillä suomenoppijoilla yhtäältä voimavara (ks. esim. Kaivapalu 2005; Kaivapalu–Eslon 2011) mutta kielten läheisyys saattaa toisaalta synnyttää sellaista varjostamista, jota ei esiinny muilla oppijoilla (ks. esim. Spoelman 2013).

Aiemmin opittujen kielten vaikutus kielenoppimiseen on läpi toisen kielen oppimisen tutkimuksen historian ajan ollut yksi alan keskeisimpiä tutkimuskohteita (ks. esim. Odlin 1989; Jarvis & Pavlenko 2008). Kuten Scott Jarvis (2000: 246–247) kuitenkin huomauttaa, tutkimus on pitkään ollut hyvin epäsystemaattista siltä osin, miten, missä, ja missä määrin aiempien kielten vaikutus todella näyttäytyy toisen kielen oppijoiden kielenkäytössä tai kielitaidossa – niin, että tulokset ovat pitkälti tapauskohtaisia ja ennen kaikkea tutkijan intuition perustuvia. Jarvisin (mp.) mukaan aiemman kielitaidon vaikutuksen on tutkimuksessa todettu joko vähenevän (ks. esim. Taylor B. 1975), kasvavan (ks. esim. Hyltenstam 1984), pysyvän muuttumattomana (ks. esim. Bohn & Flege 1992) tai vaihtelevan (ks. esim. Jarvis 1998) kielitaidon kasvaessa opittavassa kielessä. Jotta kieltenvälisiä vaikutussuhteita voitaisiin tutkia systemaattisesti ja jotta eri tutkimusten tuloksia voitaisiin verrata entistä paremmin, Jarvis (2000, 2010) ehdottaakin systemaattisten kriteerien käyttöönottoa vaikutusten tunnistamisessa ja oppijankielessä havaittavien piirteiden käsittelemisessä:

1. Toimivatko samaa ensikieltä puhuvat kielenoppijat keskenään samalla tavalla toista kieltä käyttäessään?
2. Toimivatko eri ensikieliä puhuvat kielenoppijat keskenään eri tavalla toista kieltä käyttäessään?
3. Onko tarkastelunalaisessa lähtökielessä ja kohdekielessä samankaltaisia piirteitä?
4. Onko muissa tarkastelunalaisissa lähtökielissä ja kohdekielessä toisistaan eriäviä piirteitä? (Jarvis 2000: 252; kohta 4 lisätty Jarvis 2010)

Näiden kriteerien lisäksi Jarvis (2012: 5–7) erottaa toisistaan kaksi keskenään erilaista kieltenvälisen vaikutusten tutkimuksessa sovellettavaa argumenttia. Kaikkien edellä kuvattujen neljän kriteerin täytyminen muodostaa vertailuun perustuvan todisteen kieltenvälisestä vaikutuksesta (engl. *comparison-based argument*), kun taas kahden

ensimmäisen kriteerin täytyminen muodostaa havaintoon perustuvan todisteen (engl. *detection-based argument*). Vertailuun perustuvaa argumenttia voidaan Jarvisin mukaan pitää hyvin vahvana todisteena vaikutuksille, mutta se saattaa kuitenkin herkästi jättää huomiotta sellaisia vaikutussuhteita, jotka ovat epäselvempiä, joissa tarkasteltavat kielet käyttäytyvät keskenään eri tavoin ainakin pintapuolisesti mutta jotka kuitenkin voidaan tunnistaa havaintoon perustuvien todisteiden avulla. Jälkimmäinen lähestymistapa voi nostaa esiin sellaisia tutkimuksen kannalta mielenkiintoisia piirteitä, jotka eivät esimerkiksi noudata aiempia havaintoja kieltenvälisistä samankaltaisuuksista tai jotka koskevat laajempien samankaltaisuuksien sijaan yksittäisiä kielenpiirteitä tai konstruktioita. Näkemys on hyvin relevantti tämän väitöskirjan puitteissa, sillä yhtenä keskeisenä teoreettis-metodologisena tavoitteena on nostaa esiin sellaisia oppijankielen ilmiöitä, jotka helposti jäävät havaitsematta intuitioon ja tutkijan subjektiivisiin havaintoihin tai a priori -tietoon nojaavissa tutkimusasetelmissä. Tämä puolestaan on hyvin relevanttia, kun kieltä lähestytään tämän tutkimuksen tavoin käyttöpohjaisesti ja kompleksisuusteorian kaltaisen, monitasoiset ja -tahoiset vaikutussuhteet huomioonottavan teoreettisen kehyksen näkökulmasta.

Oppijan kielitaustan lisäksi toisen kielen oppimiseen voi vaikuttaa moni muukin oppijakohtainen muuttuja (engl. *individual learner differences*). Esimerkiksi Dörnyei (2005: 5) luonnehtii oppijakohtaisiksi muuttujiksi kaikkia sellaisia tekijöitä, jotka vaikuttavat jokaiseen oppijaan ja joissa ihmiset eroavat toisistaan asteittain. Yksilökohtaisia muuttujia käsittelevässä tutkimuksessa ollaan kutakuinkin yksimielisiä siitä, että tällaisia tekijöitä ovat ainakin yksilön kielelliset kyvyt (engl. *language aptitude*), motivaatio oppimiseen (engl. *motivation*), persoonallisuus (engl. *personality*) sekä oppimisen tai oppimistilanteen synnyttämä ahdistus ja huoli (engl. *anxiety*) (ks. esim. Skehan 1989; Robinson 2002; Dörnyei & Skehan 2003; Ellis, R. 2004; Dörnyei 2005).

On syytä muistaa, että oppijakohtaiset tekijät ovat toisen kielen oppimisen keskiössä, ja niiden onkin todistettu korreloivan voimakkaasti oppimisen kanssa, jopa paremmin kuin minkään muun yksittäisen muuttujan (ks. esim. Dörnyei 2005). Tässä tutkimuksessa pyrin kuitenkin tutkimaan nimenomaan edistynyttä oppijansuomea ja sen eroja ensikielenä käytetystä suomesta. Informanttien opiskeluhistoriat ja oppimiskontekstit vaihtelevat runsaasti, mistä syystä heidän kielitaitonsa luonne ja kielitaidon eri osa-alueet saattavat vaihdella runsaastikin. Tässä tutkimuksessa lähdetään kielitaidon osalta siitä funktionaalista lähtöoletuksesta, että koska informantit opiskelevat suomalaisen yliopiston

suomenkielisessä opinto-ohjelmassa, on heidän kielitaitonsa oletettavasti kyllin edistynyttä tällaisessa kielellisesti vaativassa kontekstissa toimimiseen – oppijakohtaisista eroista riippumatta. Toisaalta tutkimuksen yhtenä päätavoitteena on löytää sellaisia oppijansuomen rakenteellisia piirteitä, jotka yhdistävät ainakin valtaosaa tutkittavista suomenoppijoista ja erottavat näitä samanaikaisesti vertailuaineiston ensikielisistä suomenkirjoittajista. Näin ollen myös tutkimuksen tulokset voivat tuoda lisävaloa sellaisiin suomen kielen rakenteellisiin piirteisiin, joiden kontekstuaalisesti konventionaalisen käytön voi ajatella indikoivan edistynyttä kielitaitoa tutkitussa kielenkäyttökontekstissa.

Tässä tutkimuksessa oppijakohtaiset muuttujat on tärkeä ottaa huomioon etenkin tulkittaessa tuloksia. Verraten suuren aineiston ja määrällisen menetelmän avulla tulokset kuvaavat tyypillistä oppijaa mutta eivät kuitenkaan välttämättä yhtäkään yksittäistä oppijaa. Dedusoidessa suurta joukkoa kuvaavia tyypillisyyksiä yksittäisiin kielenkäyttäjiin on noudatettava äärimmäistä harkintaa ja otettava tarkasti huomioon myös yksilökohtaisten muuttujien suuri merkitys. Kuten Larsen-Freeman toteaa, yksittäisiä tekijöitä on lopulta aina liikaa, joten jokaista niistä ei voi koskaan ottaa kattavasti huomioon (Larsen-Freeman 1997: 141). Tulokset voivat kertoa verraten luotettavasti, millaista on tutkimuksen kohteena oleva oppijansuomi yleisesti ja miten se tyypillisesti muuttuu. Kuten korpustutkimuksessa ylipäättäänkin, tulokset eivät kuitenkaan välttämättä kuvaa kunkin yksittäisen suomenoppijan kieltä tai kielenoppimisprosessia. Yhtäältä tämän tutkimuksen voidaan nähdä nostavan esiin yleisiä trendejä, joita vasten yksilöiden kieltä voidaan tarkastella. Toisaalta kieltä käyttävät yksilöt ovat kielenoppimisen kiistaton perusyksikkö, ja osoittamalla eri yksilöiden kielenkäytössä ja sen muutoksessa havaittavia samankaltaisuuksia ja näiden suhdetta kielenulkoiseen ympäristöön tai genren kaltaisiin kielellis-sosiaalisiin rakennelmiin tämän tutkimuksen tulokset kuvaavat näiden perusyksiköiden keskinäisiä yhteyksiä ja perusyksiköiden suhdetta ympäristöönsä. Näin tämän tutkimuksen tulokset täydentävät osaltaan yksilölähtöistä kielenoppimisen kuvausta.

2.2.4. Tekstilaji käyttökontekstin kuvaajana

Kielenkäyttö on väistämättä aina sidoksissa siihen kontekstiin, jossa sitä tuotetaan, ja siihen tarkoitukseen, jota varten sitä tuotetaan. Kuten aiemmin todetaan (ks. 2.1.1), käyttöpohjaisen oppimiskäsityksen mukaan koko kielijärjestelmä muodostuu kielen tyypillisistä ja mahdollisista käyttötavoista. Näenkin tilannekohtaiset tekijät yhtenä keskeisenä oppijankieltä – niin kuin kaikkea muutakin kielenkäyttöä – muovaavana tekijänä. Oppijankielen

tutkimuksen piirissä on esimerkiksi havaittu, että kielenoppijoiden tuottaman kielen rakenteellinen oikeakielisyys ei takaa sen pragmaattista eli käyttökontekstikohtaista soveltuvuutta tai oikeellisuutta (ks. esim. Bardovi-Harlig & Hartford 1990, 1993; Omar 1991; Celce-Murcia ym. 1995; Bardovi-Harlig & Dörnyei 1998) – tyypillisesti niin, että kontekstuaalinen kompetenssi on rakenteellista kompetenssia heikompaa (Olshtain & Blum-Kulka 1985). Tämä seikka on keskeinen myös silloin, kun puhutaan edistyneistä kielenoppijoista ja kielitaidon edistyneisyyden määrittelemisestä. Esimerkiksi Eurooppalainen viitekehys (CEF) kuvaa kielitaitoa sen mukaan, mihin toimiin yksilö kykenee kielitaidollaan. Viitekehyksessä kielitaidon katsotaan jakautuvan kielelliseen, sosiolingvistiseen ja pragmaattiseen kielitaitoon (mts. 13–14). Edistyneen kielitaidon kannalta keskeiset tasot lienevät itsenäisen kielenkäyttäjän taitotasot (B1 ja B2) sekä taitavan kielenkäyttäjän taitotasot (C1 ja C2). Kirjoittamisen osalta näitä tasoja kuvataan viitekehyksessä seuraavasti:

B1: Pystyn kirjoittamaan yksinkertaista, yhtenäistä tekstiä aiheista, jotka ovat tuttuja tai itseäni kiinnostavia. Pystyn kirjoittamaan henkilökohtaisia kirjeitä, joissa kuvailen kokemuksia ja vaikutelmia.

B2: Pystyn kirjoittamaan selkeitä, myös yksityiskohtia sisältäviä selvityksiä hyvinkin erilaisista aiheista, jotka kiinnostavat minua. Pystyn laatimaan kirjoitelman tai raportin, jossa välitän tietoa tai esitän perusteluja jonkin tietyn näkökannan puolesta tai sitä vastaan. Pystyn kirjoittamaan kirjeitä, joissa korostan tapahtumien tai kokemusten henkilökohtaista merkitystä.

C1: Pystyn ilmaisemaan ajatuksiani ja näkökantojani varsin laajasti selkeässä, hyvin rakennetussa tekstissä. Pystyn kirjoittamaan yksityiskohtaisia selvityksiä monipolvisista aiheista kirjeessä, esseessä/kirjoitelmassa tai raportissa sekä korostamaan tärkeimpinä pitämiäni seikkoja. Osaan valita oletetulle lukijalle sopivan tyylin.

C2: Osaan kirjoittaa selkeää, sujuvaa tekstiä asiaankuuluvalla tyyllillä. Pystyn kirjoittamaan monimutkaisia kirjeitä, raportteja tai artikkeleita, jotka esittelevät jonkin yksittäisen tapauksen. Käytän tehokkaasti loogisia rakenteita, jotka auttavat vastaanottajaa löytämään ja muistamaan keskeiset seikat. Pystyn kirjoittamaan koosteita ja katsauksia ammattiin tai kirjallisuuteen liittyvistä julkaisuista.

(CEF: 50–51.)

Näiden kuvausten valossa voitaneen sanoa, että itsenäisen kielenkäyttäjän ylemmän taitotason (B2) saavuttaneen oppijan lingvistiset kyvyt riittävät ilmaisemaan ajatuksia laajasti, kun taas käyttökontekstin tyypilliset piirteet huomioonottava kielitaito muotoutuu vasta taitavan kielenkäyttäjän (C1 ja C2) taitotasoilla.

Kun puhutaan kielenkäytön kontekstuaalisesta tyypillisyydestä, viitataan usein tekstin (kirjoitetun tai puhutun) genreen eli siihen tekstilajiin, jota teksti edustaa tai pyrkii edustamaan – ajatuksena on siis karkeasti ottaen hahmottaa, minkälaisen tekstien joukkoon jokin yksittäinen teksti kuuluu, jotta voimme soveltaa aiempia kokemuksiamme ja aiempaa

tietoamme tekstin ymmärtämiseen. Tekstilajin luonnetta ja tekstilaji-käsitettä on käsitelty myös Suomessa runsaasti niin lingvistiikan, kirjallisuudentutkimuksen, sosiologian kuin mediatutkimuksenkin piirissä (ks. esim. Ridell 1994; Kalliokoski & Mäntynen 2004; Hakulinen & Leino 2006; Shore & Solin 2006; Heikkinen ym. toim. 2012a; S2-alalla ks. esim. Kalliokoski 1996, 2006). En pyri tämän tutkimuksen puitteissa ilmiön laajaan kuvaukseen, sillä se on luonteeltaan hyvin moniulotteinen, minkä lisäksi tekstilajille esitetyt määritelmät poikkeavat toisistaan usein sen mukaan, minkälaiseen tutkimukseen niitä sovelletaan ja mitä tekstilajia määrittäviä seikkoja halutaan painottaa. Toisaalta käyttämäni tutkimusaineisto – edistyneiden suomenoppijoiden kirjoitetut suomenkieliset tenttivastaukset – on luonteeltaan hyvin vahvasti käyttötilanteen määrittämää, ja kontekstin, käyttötarkoituksen ja kommunikaation osanottajien välisen suhteen vaikutus teksteihin on keskeisessä asemassa saamieni tulosten tulkinnessa (ks. erityisesti *Mahdollisuuden ilmaukset* ja *Pitkittäismuutokset*). Pyrin tässä alaluvussa katsauksenomaisesti esittelemään tekstilajin merkitystä ja tekstilajikäsitteen soveltamista toisen kielen tutkimuksen piirissä niiltä osin, kuin se näyttäytyy relevanttina tämän tutkimuksen puitteissa.

Kuten esimerkiksi Päivi Valtonen (2012: 45–52) toteaa, tyypillisimpiä tekstilajin kielitieteellisen määritelmien eroja on se, lähestytäänkö tekstilajia toimintana, näin painottaen sen viestinnällistä luonnetta tietyssä tilanteessa, vai rakenteena, jolloin tekstin rakenteelliset ratkaisut ja sen suhde prototyyppiseen oman lajinsa edustajaan ovat tarkastelun keskiössä. Sunny Hyonin (1996) mukaan genreä eli tekstilajia on lähestytty kielitieteellisen tutkimuksen parissa ennen kaikkea kolmesta eri näkökulmasta (ks. vastaavasta jaosta myös esim. Hyland 2004: 24–25). Kuten todettua erot ovat monelta osin painotus- ja fokuskohtaisia, ja kuvausten eroja korostavasta luonteesta huolimatta näkökulmat lankeavat monilta osin yhteen niin, että näkökulmat voikin nähdä ennen kaikkea erilaisena sijoittumisena tekstilajia määrittävien käsitteiden jatkumomaisilla akseleilla. Niin sanottu systeemis-funktionaalinen koulukunta (engl. *systemic functional linguistics*, ks. yhteenvetona esim. Halliday 2004) korostaa tekstilajien rakenteellista olemusta eli sitä, minkälaiset kielelliset rakenteet, tekstin skemaattiset ratkaisut tai semanttiset suhteet ja konventiot ovat tyypillisiä minkäkinlaisille teksteille (ks. esim. Martin, J. 1985; Halliday & Hasan 1989). Pohjoisamerikkalainen uuden retoriikan koulukunta (engl. *new rhetorics*, ks. esim. Miller 1984; Devitt 2004) puolestaan painottaa genren kommunikatiivista luonnetta toimintana tietyssä kieliyhteisössä, ja se nojaa täten bahtinilaiseen (ks. Bakhtin 1986) ajatukseen tekstilajien dialogisuudesta, variaatiosta ja funktiosta sosiaalisten roolien

määrittäjinä ja vahvistajina. Kolmantena, muita soveltamislähtöisempänä näkökulmana näyttäytyy erityisesti akateemisissa konteksteissa käytetyn englannin tutkimukseen paneutunut, tekstien alakohtaisia erityispiirteitä korostava koulukunta (engl. *English for specific purposes*, ks. esim. Swales 1990; Bhatia 1993; Biber 2006), jonka piirissä myös oppijankielen näkökulma on vahvasti läsnä (ks. esim. Belcher & Braine toim. 1995; Tang toim. 2012).

Tekstilajitietoisuus ja sen oppiminen ovat keskeisimpiä syitä siihen, että tilannekohtaisilla tekijöillä on merkitystä toisen kielen oppimisen ja S2-tutkimuksen kannalta. Berkenkotterin ja Huckinin (1995: ix) mukaan tekstilajitietoisuus kuvaa yksilön kykyä ja valmiuksia tuottaa ja ymmärtää asiaankuuluvia reaktioita samankaltaisina toistuviin tilanteisiin. Hyland (2004: 54) toteaa tekstilajitietoisuuden olevan tärkeää toisen kielen opetukselle ja tutkimukselle, sillä tekstilajit ovat kulttuuri- ja kieliyhteisökohtaisia, mistä syystä ne saattavat erottaa ensimmäisen ja toisen kielen käyttäjiä toisistaan. Tekstilajitietoisuus on luonteeltaan abstraktia tietoa, mutta se on kehkeytynyt todellisten tekstien välityksellä käyttöpohjaisesti – kohtaamalla ja havaitsemalla analogisesti tekstien välisiä eroja ja samankaltaisuuksia. Kielenkäyttö ja kielenoppiminen ovat siis luonteeltaan kontekstikohtaisia ilmiöitä – aivan niin kuin edellä (ks. 2.1.1) esitelty käyttöpohjainen malli olettaa. Toisen kielen oppijoiden kokemuspohja kohdekielen (tässä tutkimuksessa suomi) teksteistä poikkeaa ensikielisistä, ja heillä on toisaalta kokemuspohjaa samankaltaisista teksteistä erilaisissa kulttuurikonteksteissa. Näin ollen onkin ymmärrettävää, että toisen kielen oppijoiden tekstilajitietoisuus, sen tuottamat abstraktit tekstiskeemat ja niiden avulla tuotetut tekstit voivat poiketa ensikielisten teksteistä.

Kuten aiemmin todettiin, tekstilajilla on erilaisista näkökulmista ja koulukuntien välisistä eroista riippumatta runsaasti sellaisia ominaisuuksia, jotka määrittävät sen luonnetta teoreettisena käsitteenä ja joiden kautta sitä voidaan lähestyä myös toisen kielen oppimisen tutkimuksessa ja opetuksessa. Hyland (2004) tiivistää tekstilajin toisen kielen oppimisen ja opetuksen kannalta keskeiset tekijät seuraavaan kahdeksaan, teoreettisesta näkökulmasta riippumattomaan kohtaan:

1. Tekstilajit kehittyvät niiden toistuvien tapojen seurauksena, joiden avulla ihmiset hoitavat asioita sosiaalisessa ryhmässään;
2. Nämä ryhmät ovat verrattain stabiileja, ja instituutioiden muodostamat tekstilajit stabiloituvat ajan kuluessa tiettyyn pisteeseen asti, mikä puolestaan edistää ymmärtämistä ja koherenssia;

3. Tekstilajeilla on eriteltävissä olevia kielellisiä piirteitä, jotka eivät ole täysin kontekstin tai tekstilajin määriteltävissä mutta eivät myöskään täysin yksittäisten tuottajien ratkaisujen varassa;
 4. Tekstit eivät ole ainoastaan itseään ilmaisevien yksilöiden tuottamia, vaan ympäröivä yhteisö ja kulttuuri vaikuttavat sekä lopputulokseen että itse tuottamisprosessiin. Tästä syystä tekstit muuttuvat muuttuvien tarpeiden mukaan;
 5. Tekstilajin ymmärtäminen vaatii sekä muodon että sisällön ymmärtämistä sekä sen käsittämistä, mikä on luonteenomaista kulloisellekin tavoitteelle ja kontekstille;
 6. Tekstien kieltä tulisi opettaa yhdessä niiden tyyppillisen käyttöympäristöjen kanssa;
 7. Tekstilajeilla on sosiaalinen lähtökohta, joten eri tekstilajit kantavat muassaan taustan mukanaan tuomia valta- ja statusrakenteita;
 8. Kaiken kirjoituksen opetuksen tulisi sisältää tekstien tyyppillisten piirteiden ja niihin liittyvän sosiaalisen vallan käsittelyä.
- (Hyland 2004: 51.)

Niin ikään Valtonen (2012: 47) huomauttaa, että erilaiset genren määritelmät eivät suinkaan sulje toisiaan pois, vaan ne voi nähdä toisiaan täydentävinä siten, että tekstilajia voi lähestyä samanaikaisesti useasta eri näkökulmasta. Tämä tuntuu mielekkäältä myös toisen kielen tutkimuksen viitekehyksessä. Yhtäältä on relevanttia lähestyä tekstejä Hallidayta ja Hasania (1976) seuraten tarkastelemalla sitä sosiaalista asetelmaa (engl. *field*), johon osaaottavat kielenkäyttäjät kuuluvat ja johon teksti sitoutuu, kommunikaation osanottajien välisiä suhteita (engl. *tenor*) sekä käytetyn kielen roolia, tuottotapaa ja esityskontekstia (engl. *mode*). Näitä muuttujia määrittävät tekstin kontekstuaalinen funktio, tuottajan tavoite ja mahdollisuus valmistautumiseen, tuotantotapa sekä kommunikaatioon osallistujien sosiaaliset suhteet (mp.). Ensimmäisen ja toisen kielen käyttäjien välisten erojen on helppo nähdä vaikuttavan näihin kaikkiin tekijöihin – sosiaalinen asetelma ja tavoitteet poikkeavat usein toisistaan, samoin osanottajien väliset suhteet ja kielen rooli kulloinkin käsillä olevassa kommunikaatiossa. Toisen kielen käyttäjän yhtenä määrittävänä tekijänä sosiaalisessa asetelmassa on usein se, että hän on tietyllä tavalla vieras, ulkopuolinen tai tulija, mikä puolestaan väistämättä vaikuttaa hänen suhteeseensa muihin kielenkäyttäjiin ja siihen, millaisessa roolissa hänen käyttämänsä kieli nähdään osana viestintää.

Vastaavasti John Swalesin (1990: 58) käyttötilannelähtöinen näkemys tuntuu perustellulta myös toisen kielen oppimisen viitekehyksessä: Swalesin (mp.) mukaan tekstilajit ovat kommunikaatiotapahtumia, joiden osanottajat jakavat tietyt viestinnälliset tavoitteet, jotka puolestaan ovat kulloisenkin viestijäyhteisön – kuten tietyn tieteenalan tutkijoiden ja opettajien – hyväksymiä. Swales lähestyy määritelmässään myös rakenteellista näkökulmaa todetessaan tietyn tekstilajin edustajien olevan yhteisen tavoitteen lisäksi samankaltaisia niin rakenteen, tyylin, sisällön kuin oletettujen vastaanottajienkin osalta. Kun tuottajana on toisen kielen oppija, voivat viestinnälliset tavoitteet kuitenkin olla erilaiset esimerkiksi siksi, että

tuottajan asema viestijäyhteisössä on erilainen, tai siksi, että hän ei tavoitellun tekstilajin tunnistamisesta huolimatta tunne niitä rakenteellisia seikkoja, joita tavoitellulta tekstilajilta odotetaan.

Jan-Ola Östman (2005) pyrkii diskurssikaava-käsitteen (engl. *discourse pattern*) avulla jäsentämään kielen käyttöympäristön ja sen kieliopillisten yksiköiden välistä suhdetta konstruktiokieliopin näkökulmasta. Östmanin (mp.) mukaan diskurssikaavat ovat konstruktioiden kaltaisia mutta niitä laajempia abstraktioita, joissa teksteille tyypilliset piirteet ovat assosioituneet niille tyypillisiin käyttöympäristöihin. Nämä diskurssikaavat muodostavat kielenkäyttäjille kehyksiä, joiden puitteissa tekstejä tulkitaan ja niitä ymmärretään. Östman korostaa diskurssikaavan olevan konstruktion tavoin kognitiivinen abstraktio, ja käyttö pohjaisen oppimiskäsityksen hengessä onkin luontevaa ajatella tämän abstraktion perustuvan aiemmassa syötöksessä havaittuihin piirteiden ja niiden käyttöympäristöjen yhteisesiintymiin. Diskurssikaavoja määrittävät tekijät lienevät monilta osin juuri Swalesin (1990: 58) mainitsemia tekstilajia määrittäviä seikkoja: tekstin viestinnällinen tavoite, rakenne, tyyli, sisältö ja sen oletetut vastaanottajat. Tästä näkökulmasta onkin helppo ymmärtää diskurssikaavojen soveltamisen erottavan ensikielisiä kielenkäyttäjiä ensikielisistä – puhtaasti siksi, että jotkin ensikielisille kielenkäyttäjille tutut käyttöympäristöt saattavat olla toisen kielen käyttäjille vieraita tai tietyille käyttöympäristöille tyypilliset diskurssikaavat saattavat olla eri kielenkäyttäjille erilaisia yksilö- tai ensikielikohtaisesti erilaisen aiemman syötöksen vuoksi.

Tämä tutkimus käsittelee suomea toisena kielenä käyttävien kirjoittajien suomenkielisiä yliopistossa kirjoitettujen tenttien tenttivastauksia. Käytän tässä väitöskirjatutkimuksessa käsitettä tekstilaji (englanninkielisissä artikkeleissa *genre*) kuvaamaan niiden tekstien luonnetta, joita tutkimusaineistoni tekstit edustavat. Tekstit ovat kirjoitettua akateemista kieltä, ja ne edustavat sen alalajia tenttivastausta. Tekstit ovat siis luonteeltaan akateemisia (ne liittyvät sisällöllisesti kirjoittajien yliopisto-opintoihin), niillä on hyvin erikoistunut käyttöympäristö (ne pyrkivät osoittamaan kirjoittajan tiedot tietystä tarkkaan rajatusta aiheesta, kuten tietty kurssi tai tietyt opiskelumateriaalit), tekstien tuottamistilanne on tarkkaan rajattu (ne on kirjoitettu valvotussa tenttitilaisuudessa) ja kommunikaation ensisijaiset osanottajat ja heidän ensisijaiset roolinsa ovat hyvin tiedossa (tekstien kirjoittajat ovat suomea ensimmäisenä tai toisena kielenä käyttäviä yliopisto-opiskelijoita, jotka tekevät yliopisto-opintoja suomesta tai suomalais-ugrilaisesta kielentutkimuksesta laajasti ajateltuna;

tekstien ensisijaiset lukijat toimivat opetustehtävissä suomalaisessa yliopistossa suomen kielen tai suomalais-ugrilaisen kielentutkimuksen alalla). Tehtävänannot kuitenkin vaihtelevat tekstien välillä runsaasti niin luonteeltaan kuin rakenteeltaankin (esimerkkejä tämän tutkimusaineiston tehtävänannoista, ks. *Pitkittäismuutokset*; tehtävänannoista yleisesti ks. esim. Makkonen-Craig 2008), mikä puolestaan voi vaikuttaa tuotetun kielen luonteeseen. Esimerkiksi Lia Plakans ja Atta Gebril (2013) huomauttavat, että kielenkäyttäjän kielitaitotaso voi vaikuttaa tehtävänannon tai lähdetekstin käyttöön siten, että edistyneemmät englanninoppijat käyttävät alkeisoppijoita vähemmän lähdetekstin sanastoa. Toisaalta Outi Toropainen (2014) huomauttaa, että tällaista ei ole ainakaan yksiselitteisesti nähtävissä S2-kielenkäyttäjien teksteissä.

Ajatukseni tekstilajista on siis hyvin käyttö- ja funktiolähtöinen, ja se vastaa monilta osin edellä kuvattua, erikoiskielten tutkimuksessa sovellettua ajatusta tekstilajista joukkona rakenteeltaan osin samankaltaisia kommunikatiivisia toimia. Tutkimani tekstit ja niiden kautta käytävään kommunikaatioon osallistuvat ihmiset jakavat ainakin laajasti ajateltuna tietyt sosiaaliset tavoitteet (ks. esim. Swales 1990: 45–47), jotka osaltaan määrittävät käsiteltävää tekstilajia. Tutkimukseni on määrällinen ja rakennelähtöinen, sillä tarkastelen sellaisia kielellisiä piirteitä, joiden käyttö verraten suuren korpusaineiston tarkastelun perusteella tyypillisesti erottaa suomea ensikielenä ja toisena kielenä käyttäviä kirjoittajia toisistaan (ks. *S2-avainrakenteet* ja *Mahdollisuuden ilmaukset*), kielitaustaltaan erilaisia, suomea toisena kielenä käyttäviä kirjoittajia toisistaan (ks. *Siirtovaikutus-avainrakenteet*) tai joiden käyttö muuttuu suomea toisena kielenä kirjoittavien teksteissä ajan kuluessa (ks. *Pitkittäismuutokset*). Näin ollen prototyypinen tutkimani tekstilajin edustaja määrittyy tutkimusaineistossani esiintyvän vaihtelun puitteissa. Paneudun tutkimuksessani erityisesti niihin piirteisiin, joita käytetään aineiston tarkastelun perusteella usein tai toistuvasti eri tavalla aineistoni eri osissa. Tällaisten tekijöiden paikantamisen avulla päästään aineistolähtöisesti käsiksi tämän tutkimuksen perimmäiseen kysymykseen: niihin kielellisiin ja kielenkäyttöön liittyviin seikkoihin – ja edelleen mahdollisesti niiden syihin – jotka erottavat vielä edistyneitäkin toisen kielen oppijoita ensikielisistä kielenkäyttäjistä.

3. AVAINRAKENNEANALYYSI KORPUSTUTKIMUKSEN KENTÄLLÄ

Olen asettanut tälle väitöskirjatutkimukselle kaksi rinnakkaista ja toisiaan tukevaa tavoitetta: yhtäältä tarkastella ja kehittää sellaisia korpustutkimuksen menetelmiä, joiden avulla oppijankielelle tyypillisiä rakenteellisia seikkoja voidaan tarkastella aineistovetoisesti – niin, että tarkasteltavia rakenteita ei ole ennalta määritetty – ja toisaalta soveltaa näitä menetelmiä edistyneiden S2-oppijoiden akateemisten tekstien analyysiin ja näille teksteille tyypillisten (tai epätyypillisten) konstruktionaalisten piirteiden tarkasteluun sekä eron taustalla vaikuttavien syiden tulkintaan. Aloitan tämän luvun käsittelemällä lyhyesti oppijankielen korpusten ominais- ja erityispiirteitä (3.1). Esittelen sen jälkeen korpusvetoisuuden ja tilastollisen avaimuuden käsitteitä oppijankielen korpustutkimuksen kannalta (3.2). Päätän luvun esittelemällä niitä kontrastiivis-deskriptiivisiä tutkimusmenetelmiä, joiden varaan tämän tutkimuksen analyysi rakentuu ja pohtimalla lyhyesti menetelmällisten ratkaisujen vaikutusta tutkimukseen ja siitä saatavien tulosten luonteeseen (3.3). Tämän tutkimuksen aineiston ja käytetyt menetelmät esittelen erikseen luvussa 4.

3.1. Korpus oppijankielen tutkimusaineistona

Kielikorpuksella tarkoitetaan laajasti ottaen suurta kieliaineistoa, jonka avulla voidaan tarkastella sitä kielivarianttia, jota kyseessä oleva korpus edustaa. Korpuslingvistiikan keskeisimpiä pyrkimyksiä on kuvata tiettyä kieltä tai kielimuotoa (Francis 1993: 17) abstrahoidun mallin sijaan sellaisena kuin sitä todellisuudessa käytetään. Tästä syystä on ensiarvoisten tärkeää hahmottaa, mitä kulloinkin käsillä oleva korpus edustaa, sillä se määrittää korpuksen avulla saatavien tulosten yleistettävyyttä (*LAS2*; ks. myös esim. Atkins ym. 1992; Biber 1993; Sinclair 1996; Laviosa-Braithwaite 1996; Granger 2002; Hunston 2008; Jantunen 2011b). Yleisiä korpusten edustavuutta määrittäviä tekijöitä ovat tyypillisesti korpuksen ja sen osakorpusten koko, aineiston autenttisuus sekä aineiston keruutapa (Biber 1993; Sinclair 1996; Tognini-Bonelli 2001: 47–64; Hunston 2008). Edustavuutta määrittävien tekijöiden huomioimisen lisäksi korpuksiin liitetään usein monenlaisia annotaatioita eli kulloinkin käsillä olevan korpuksen kannalta mielekästä metatietoa. Tällaisia metatietoja voivat olla esimerkiksi tuottamisajan, -paikan, julkaisukanavan tai kirjoittajan kaltaiset tekstin ulkopuoliset taustamuuttujat, tekstin sisäinen jakso- kappale-, virke- tai lauserakenne tai metalingvistinen annotaatio eli esimerkiksi sanaluokkien, morfologisten muotojen tai syntaktisten funktioiden merkitseminen aineistoon (ks. esim. Garside ym. toim. 1997).

Muun korpustutkimuksen tavoin oppijankielen korpustutkimus (engl. *learner corpus research*) on lisääntynyt huomattavasti tietokoneiden kehittymisen myötä 1990- ja 2000-luvuilla (ks. yhteenvetona oppijansuomen korpustutkimuksesta esim. Jantunen & Piltonen 2009; Jantunen & Pirkola 2015). Oppijankielen korpukset lasketaan usein niin sanottujen erikoiskorpusten joukkoon. Tognini-Bonellin mukaan erikoiskorpukset eroavat perustavanlaatuisesti normaalia, autenttista kieltä kuvaavista korpuksista, sillä ne eivät ole osa jonkin luonnollisessa ympäristössä käytettävän kielivariantin neutraalia kuvausta. (Tognini-Bonelli 2001: 8–9.) Oppijankielen korpusaineistojen autenttisuuden määrittäminen on vaikeaa – aineisto on usein kerätty osana kielen opiskelua, ja niin oppimistilanne kuin kulloinkin tuottamista määrittänyt tehtävänanto vaikuttavat luonnollisesti sisältöön. Toisaalta oppijankieltä sisältävien korpusten ryhmitteleminen yhtenäiseksi joukoksi ei useinkaan vastaa todellisuutta, sillä eri korpukset eroavat usein toisistaan esimerkiksi käyttö- ja oppimiskontekstin osalta. (Granger 2002: 8.) Variaatio on huomattavaa jo eri oppijansuomea sisältävien korpusten välillä niin tuottamis- ja oppimisympäristön, kielitaitotason, sisällytettyjen tekstilajien, aineiston pitkittäisyyden kuin oppijoiden ensikielten ja iänkin osalta (Jantunen & Pirkola 2015), mistä syystä taustamuuttujien huomioonottaminen oppijankielen korpuksissa on ensiarvoisen tärkeää (LAS2; ks. myös Granger 2008: 263–265; Jantunen 2011b).

3.2. Korpusvetoisuus ja tilastollinen avaimuus

Korpukset mahdollistavat hyvin monenlaisia tutkimusasetelmia, eikä korpusten käyttö kerro asetelmasta juuri muuta kuin sen, että se perustuu ainakin joiltakin osin todelliseen kielenkäyttöön. Korpusten rooli kielentutkimuksen aineistona on puhuttanut niin kielitieteen tutkijayhteisöä ylipäätään kuin erityisesti suomen kielen tutkimuksenkin osalta (katsauksena keskusteluun ks. esim. Huumo 2007; Ojutkangas 2008; Jantunen 2009a, Brattico 2013). Yhtenä lähestymistapoja määrittävänä erona on pidetty jakoa korpuspohjaiseen (engl. *corpus-based*) ja korpusvetoiseen (engl. *corpus-driven*) lähestymistapaan sen mukaan, onko korpusten roolina ennen kaikkea tarkastella jonkin hypoteesin toteutumista tai jonkin ennalta määritellyn kielenpiirteen tyypillistä käyttöä (korpuspohjainen lähestyminen) vai käytetäänkö korpusta tämän tutkimuksen tapaan jo itse tutkimuskysymyksen ja sille relevantin tutkimusasetelman muotoilemisessa (korpusvetoinen lähestyminen) (Tognini-Bonelli 2001; Hardie & McEnery 2010; ks. myös esim. Jantunen 2009a tutkimuksen jaosta aineistoesimerkein tuettuun, aineistopohjaiseen ja aineistovetoiseen lähestymiseen).

Yksi korpusvetoisen tutkimuksen keskeisimpiä käsitteitä on avaimuus (engl. *keyness*) ja tyypillisimpiä analyysikeinoja siihen perustuva avainsana-analyysi (engl. *key word analysis*). Avainsanoja ja avaimuuden käsitettä käytetään kuitenkin tutkimuksessa monin eri tavoin (yhteenvedona ks. esim. Stubbs 2010). Suurimpina yhdistävinä tekijöinä on ajatus diskurssista toistuvina konventionaalisina kommunikoinnin tapoina (mts.: 22). Avainsanoja on pidetty esimerkiksi kulttuuristen fokalisoitipisteiden osoittimina (ks. esim. Williams 1976; Wierzbicka 1997), tilastollisesti poikkeavasti tietyssä aineistossa käyttäytyvinä sanoina (ks. esim. Scott–Tribble 2006) tai skeemoina, joissa samaa funktiota voi olla ilmaisemassa erilaisia lekseemejä (ks. esim. Francis 1993). Mike Scott (2010) kuvaa avaimuuden käsitteen ja avainsana-analyysin tavoitetta korpustutkimuksen kannalta seuraavasti:

”A key grants access to a place which is otherwise restricted, private, sealed off. In this sense a word or phrase is a key which enables one to see something, it is an enabling device.”
(Scott 2010: 44.)

Scott on pohtinut tilastollisen avaimuuden käsitettä runsaasti ja esittänyt käytännön työkaluja avainsanojen tavoittamiseksi (mm. Scott & Thompson 2001; Scott & Tribble 2006; Scott 2010). Scott rajaa termin koskemaan sanoja, sanaklustereita ja fraaseja (Scott 2010: 43), joiden hän katsoo olevan sellaisia leksikaalisia elementtejä, jotka ovat keskeisiä tarkasteltavana olevalle aineistolle suhteessa johonkin muuhun aineistoon. Jonkin kielenpiirteen avaimuus on siis aina tekstiriippuvaista, ja sen on perustuttava tarkasteltavan aineiston ja vertailuaineiston väliseen esiintymien frekvenssieroön (Scott & Tribble 2006: 58). Avaimuuden määrittely perustuu useimmiten tilastollisiin testeihin, minkä lisäksi esiintymien absoluuttisen määrän on ylitettävä jokin avainelementiltä vaadittava, ennalta määritetty kynnyisarvo (mts. 59). Käytännössä kyse on siis tilastollisesti merkitsevistä eroista aineistojen välillä, mutta termillä avaimuus halutaan korostaa havaittujen erojen merkitystä avaimena sillä saavutettavaan tietoon käsillä olevan aineiston erityispiirteistä. Scott puhuu leksikaalisesta tarkastelusta mutta vastaavaa ajatusta voidaan soveltaa myös lingvistisen annotaation avulla tehtävään kielen rakenteellisten erojen tarkasteluun – avaimuuden käsite ei itsessään ota kantaa avaimina toimivien kielenpiirteiden luonteeseen tai yksiköiden kokoon. Tästä syystä käytän tässä tutkimuksessa avainsana-termin sijaan termiä avainmuoto, kun taas termi avainrakene viittaa niihin konstruktioihin, jotka tarkemman analyysin perusteella vaikuttavat havaitun eron taustalla.

Scottin hahmottelema näkemys jossakin aineistossa esiintyvien kielenpiirteiden avaimuudesta suhteessa johonkin toiseen aineistoon ei suinkaan ole uusi, ja esimerkiksi

oppijankielen tutkimus on perustunut kontrastiivisen analyysin hypoteesin (ks. Lado 1957) synnyn myötä usein aineistojen väliseen vertailuun ja niistä nousevien erojen analyysiin. Oppijansuomen korpustutkimuksessa tämän väitöskirjatutkimuksen osatutkimusten ohella ennen kaikkea Jantunen on asemoinut tutkimustaan nimenomaan avainsana-analyysin ja kontrastiivisen välikieliansalyysin leikkauspisteeseen (ks. esim. Jantunen 2007; Jantunen 2009b; Jantunen 2011a). Korpusvetoisessa oppijankielen tutkimuksessa tuntuu luontevalta lähestyä oppijankielen kieliopillisia ominaispiirteitä avaimuuden näkökulmasta. Näin aineistosta voidaan ilman ennalta esitettyä hypoteesia paikantaa sen avainmuotoja. Esimerkiksi Jan Aarts ja Sylviane Granger (1998) tarkastelevat eri ensikieliä puhuvien englanninoppijoiden ja Wybo Wiersma ym. (2011) ensikielisten ja ei-ensikielisten englanninkirjoittajien tyypillisiä eroja vertaamalla sanaluokka-annotaatiosta kerättyjen 3-grammien (ks. 2.1.4) frekvenssejä eri osa-aineistoissa. Tämän väitöskirjan osatutkimusten lisäksi esimerkiksi Jantunen (2011a) soveltaa metodia leksikaalisten erityispiirteiden ohella myös lingvistisen metatiedon tarkasteluun oppijansuomen korpustutkimuksessa.

3.3. Deskriptio ja kontrastointi oppijankielen korpustutkimuksessa

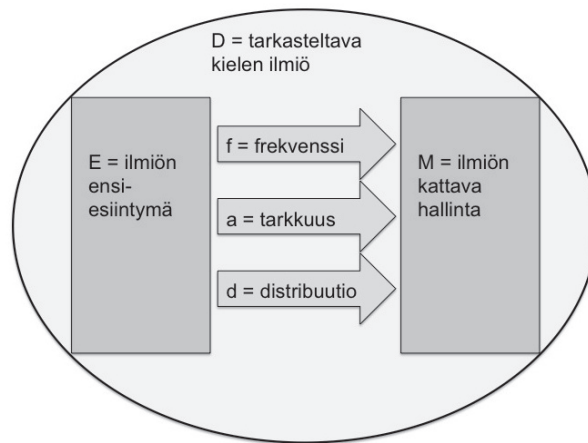
Korpustutkimuksen tarjoamien mahdollisuuksien kannalta on keskeistä huomata, että toisen kielen tutkimuksessa historiallisesti keskeisen virheanalyysin ohella korpukset mahdollistavat oppijankielen määrällisten ominaispiirteiden tarkastelun ottamatta kantaa kielen oikeellisuuteen (ks. kuitenkin Dagneaux ym. 1998 virheanalyysin soveltamisesta oppijankielen korpustutkimuksessa). Näin oppijankielen tyypillisiä piirteitä, niiden käyttöä ja jakaumia voidaan tarkastella deskriptiivisestä näkökulmasta tai vertailevasti. Vertailevassa asetelmassa ajatuksena on siis tarkastella rinnakkain joidenkin kielenpiirteiden käyttöä esimerkiksi toista kieltä käyttävien ja ensikielisten kielenkäyttäjien aineistoissa tai jonkin tietyn taustamuuttujan mukaan eroteltujen toista kieltä käyttävien kielenkäyttäjien aineistoissa. Granger (1996, 2015) kutsuu tätä menetelmää kontrastiiviseksi välikieliansalyysiksi (engl. *Contrastive Interlanguage Analysis, CIA*). Taustamuuttujia vaihtamalla vertailu voidaan kohdistaa esimerkiksi ensikielisten ja ei-ensikielisten kielenkäyttäjien välisiin eroihin (ks. *S2-avainrakenteet* ja *Mahdollisuuden ilmaukset*), eri ensikieliä puhuvien ei-ensikielisten kielenkäyttäjien välisiin eroihin (ks. *Siirtovaikutus-avainrakenteet*), eri taitotasoilla olevien ei-ensikielisten kielenkäyttäjien välisiin eroihin tai eri tekstilajien välisiin eroihin. Suomea toisena kielenä tarkastelevassa korpustutkimuksessa vertailevaa välikieliansalyysia tai sen kaltaisia vertailevia tutkimusasetelmia ovat käyttäneet tämän väitöskirjan osatutkimusten lisäksi muun muassa Marianne Spoelman (2010, 2011, 2013) tutkiessaan ensikielten vaikutuksia oppijansuomen partitiivin käytössä, Jantunen

(2011a) etsiessään tyypillisiä tai epätyypillisiä elementtejä ensikielisiin teksteihin verrattaessa, Annekatrin Kaivapalu ja Pille Eslon (2011) tarkastellessaan lähisukukielten vaikutuksien symmetrisyyttä suomen ja viron välillä sekä Jarmo Jantunen ja Sisko Brunni (2012) analysoidessaan morfologista virittymistä oppijansuomessa.

Kontrastiivinen välikieliansalyysi tarjoaa luontevan näkökulman kielen käyttöpohjaiseen olemukseen, sillä sen avulla on mahdollista kuvata tarkasteltavien kielenpiirteiden tyypillisessä käytössä havaittavia eroja suhteessa niiden funktionaaliseen variaatioon. Avainmuotojen tilastollinen paikantaminen puolestaan mahdollistaa ilmiön lähestymisen korpusvetoisesti. Tilastollinen merkitsevyys ei kuitenkaan vielä itsessään tarkoita käytännön merkitystä, minkä lisäksi käytännössä merkittävät ilmiöt eivät aina ole tilastollisesti merkitseviä – tilastollisten testien tulosten tulkintavastuu on luonnollisesti tutkijalla. Monesti oppijankielen korpustutkimuksen piirissä tehdyt tulkinnat aineistossa havaituista tyypillisistä eroista jäävät kuitenkin verraten karkeiden, kokonaisfrekvenssejä tarkastelevien tilastollisten testien avulla havaittujen tilastollisten erojen toteamiseen tai havaittuun eroon perustuvien suorien laadullisten päätelmien tekoon niin, että aineiston sisäistä variaatiota, käyttöympäristön vaikutusta tai mitään muita kielellisiä tai muita taustamuuttujia ei problematisoida niiden ansaitsemalla tavalla tai oteta lainkaan huomioon (kritiikkiä ovat esittäneet esim. Jensen 2013; Gries & Deshors 2014; Gries & Adelman 2014). Teoreettiselta kannalta kritiikki lähenee Larsen-Freemanin ja Lynne Cameronin (2008: 133) konstruktiokielioppia kohtaan esittämää kritiikkiä siitä, että se keskittyy liiaksi yksittäisiin konstruktioihin ottamatta huomioon konstruktioiden suhdetta niitä ympäröivään kielijärjestelmään ja kielenulkoiseen maailmaan. Metodologian kannalta yksi keskeisistä tekijöistä on valita tarkasteltavat yksiköt ja tilastolliset menetelmät siten, että niiden avulla kyetään ottamaan huomioon myös aineiston sisäinen vaihtelu, mahdollinen aineiston sisäinen epätasapaino ja niiden vaikutukset aineistojen väliseen vertailuun (ks. 4.2).

Vertaileva tutkimusasetelma ei suinkaan ole ainoa oppijankielen korpusten mahdollistama lähtökohta, ja esimerkiksi jonkin kielenilmiön käytön ja distribuution muutosten tarkastelu deskriptiivisesti eri taitotasoilla tai pitkittäistarkastelussa tarjoaa mahdollisuuden tarkastella oppijankielen tyypillisiä piirteitä tai oppimisen kehityskulkuja. Martin ym. (2010) kuvaavat ennalta määritettyjen konstruktioiden tai konstruktiojoukkojen (paikallissijat, transitiivikonstruktio ja passiivi) oppimista DEMfad-malliksi kutsumansa menetelmän avulla (ks. kuvio 2). Heidän mukaansa tarkasteltavan kielenpiirteen ($D = domain$) oppimista voidaan

käyttöpohjaiseen näkemykseen nojaten tarkastella konstruktion ensiesiintymisen (E = emergence)⁴ ja sen kattavan hallinnan (M = *mastery*) välillä etenevänä prosessina. Ensiesiintymisen ja kattavan hallinnan välistä jatkumomaista prosessia voidaan tarkastella ensi sijassa konstruktion käyttömäärässä (f = *frequency*), käytön tarkkuudessa (a = *accuracy*) ja käytön funktionaaliskontekstuaalisessa jakaumassa (d = *distribution*) havaittavien muutosten avulla. Martinin ym. mukaan malli sinänsä ei linkity mihinkään tiettyyn teoreettiseen asetelmaan, mutta he soveltavat sitä itse käyttöpohjaiseen kielikäsitteeseen nojaavalla tavalla, ja tarkasteltavat ilmiöt on määritelty konstruktiokieliopin mukaisina konstruktiona. (Martin ym. 2010: 58–62.)



Kuvio 2. Yksittäisen kielen ilmiön oppimisprosessia kuvaava DEMfad-malli (Martin ym. 2010: 59).

Deskriptiiviseen tutkimusasetelmaan nojaavaa oppijansuomen korpustutkimusta on tehty tämän väitöskirjan osatutkimusten ohella S2-tutkimuksen parissa runsaasti niin eksplisiittisesti edellä kuvattuun DEMfad-malliin niveltyn kuin muutenkin. Taitotasojen välisiä eroja ja niiden indikaattoreita käsittelevät esimerkiksi Reiman (2011, 2014) tarkastellessaan transitiivikonstruktion kehitystä, Kirsti Siitonen ja Jenny Niemelä (2011) edistyneiden suomenoppijoiden pitkittäismuutoksia ja Kirsti Siitonen ja Maisa Martin (2012) *U*-verbejä käsittelevissä tutkimuksissaan, Seilonen (2013) tutkiessaan epäsuoria henkilöön viittaamisen keinoja, Kajander (2013) eksistentiaalikonstruktiota käsittelevässä tutkimuksessaan sekä Keaty Siivelt ja Sanna Mustonen (2013) tarkastellessaan paikallissijojen kehitystä. Yleisempiä ilmiökohtaisia tarkasteluja ovat puolestaan tehneet

⁴ Käytän tässä tutkimuksessa englannin *emerge*-verbin suomenkielisenä vastineena verbiä *kehkeytyä*, joka kuvaa kieliopin tai jonkin yksittäisen konstruktion käyttöpohjaista vähittäistä hahmottumisprosessia. Nähdäkseni sana *emergence* kuvaa DEMfad-mallissa kuitenkin prosessin sijaan konkreettista ensiesiintymistä, mistä syystä en käytä tässä yhteydessä termiä *kehkeytyminen*.

muun muassa Tuija Määttä (2011, 2013, 2014) ruotsinkielisten suomenoppijoiden paikallissijoja ja kommunikaatioverbejä käsittelevissä tutkimuksissaan sekä Tanja Seppälä (2013) tarkastellessaan oppijansuomen ketjuuntuvia verbirakenteita.

Scott toteaa avainsanojen suurimman annin olevan osoittimina toimiminen sellaisiin kohtiin, joihin tarkempi analyysi kannattaa kohdistaa (Scott 2010: 56–57). Toinen tärkeä seikka on kulloinkin tarkasteltavan kielenpiirteen todellisen käytön havainnoiminen niin siinä esiintyvän sisäisen vaihtelun, sen tyypillisten käyttöympäristöjen kuin sen funktionaalisen distribuutionkin osalta. Osaltaan tähän voidaan vastata juuri edellä esitetyn DEMfad-mallin avulla – frekvenssin lisäksi formaalinen ja funktionaalinen distribuutio sekä laadullinen ja määrällinen kohdekielisyys ovat kiistämättä huomioon otettavia seikkoja. DEMfad-mallin avulla tehtävä tarkastelu puolestaan vastaa monilta osin Gill Francisin (1993) esittämiä metodologisia suuntaviivoja aineistovetoisen kieliopin kuvauksen tekoon. Francisin mukaan määrittely perustuu ennen kaikkea kielellisessä yksikössä esiintyvän sisäisen variaation sekä sen tyypillisessä käyttöympäristössä ja tyypillisissä funktioissa esiintyvän variaation tarkasteluun ja kuvaukseen (engl. *analysis of item and its environment*), ja näin kielenpiirteiden – kuten konstruktoiden – muodon ja merkityksen välistä suhdetta voidaan kuvata niiden tyypillisen käytön kannalta.

Vertaileva ja kuvaileva tutkimusasetelma eivät ole millään lailla vaihtoehtoisia lähestymistapoja, ja esimerkiksi korpusten avulla tehtävä vertailu on luonteeltaan nimenomaan deskriptiivistä. Näin on laita myös valtaosassa edellä mainittuja S2-tutkimuksia. Tutkimusasetelmat ja -aineistot ovat kuitenkin usein osin erilaisia, sillä yksistään kuvailevissa tutkimuksissa tarkastelunalaista kielenilmiötä havainnoidaan vain yhtä aineistoa käyttäen, kun taas vertailevissa tutkimuksissa kuvataan ilmiötä kahdessa eri aineistossa tai jonkin taustamuuttujan mukaan erotetuissa aineiston osissa ja pohditaan kuvailevan analyysin perusteella aineistojen välisiä eroja. Ensikielisiä ja ei-ensikielisiä kielenkäyttäjiä vertailevaa asetelmaa on kritisoitu oppijankielen tutkimuksessa ylipäätään – usein aiheellisesti – sen tuottaman vertailuharhan (engl. *comparative fallacy*) vuoksi siksi, että oppijankieli nähdään kohdekielen epätäydellisenä edustajana eikä kielenoppijan ja oppijan välikielen lähtökohdista (Bley-Vroman 1983, ks. myös Selinker 1972). Toisaalta kontrastiivisen välikielianaalyysin avulla voidaan verrata myös ensikielikohtaisia eroja oppijankielessä (Granger 1996, ks. myös *Siirtovaikutus-avainrakenteet*). Lisäksi ainakin aineistolähtöisesti on vaikea tarkastella sitä, mikä on ominaista, tyypillistä tai erityistä juuri oppijankielelle, ellei sitä verrata mihinkään

muuhun aineistoon (Osborne 2013). Jonkin kielimuodon tyypillisyydet voidaan tavoittaa aineistosta nimenomaan vertaamalla sitä johonkin toiseen, ja vertailu kuvaa ensi sijassa juuri näiden aineistojen välistä suhdetta – ei esimerkiksi tutkittavan oppijankielen suhdetta kaikkeen muuhun kieleen. Tutkimus- että vertailuaineiston rinnakkainen kattava kuvaileva analyysi varmistaa sen, että tutkimus todella ottaa huomioon kulloinkin käsiteltävän kielenilmiön todellisen variaation, rikkauden ja tyypillisyydet niin, että se ei nojaa eksplisiittisesti tai implisiittisesti ainoastaan tutkijan omaan intuitioon – tähän pyrin myös tässä tutkimuksessa (ks. luku 4 sekä *S2-avainrakenteet*, *Siirtovaikutus-avainrakenteet*, *Pitkittäismuutokset* ja *Mahdollisuuden ilmaukset*).

On selvää, että etenkin kielenulkoisten muuttujien huomioonottaminen korpusaineiston avulla ei aina ole kovinkaan helppoa. Kyse on osin myös tutkimusten toteutuskelpoisuudesta – mitä enemmän ennalta tarkkaan määritettyjä muuttujia halutaan ottaa huomioon, sitä rajatumppia ovat käyttökelpoiset aineistot ja sitä tarkemmin ennalta rajattu on tutkimuskysymyksen oltava, jotta tutkimusasetelmasta tulee edes jossain määrin hallittavissa oleva kokonaisuus. Tämän väitöskirjatutkimuksen kokonaislähestyminen on leimallisesti aineisto- eli korpusvetoinen, sillä yhtenä keskeisenä kysymyksenasetteluna on se, miten oppijankielelle tyypillisiä piirteitä voidaan lähestyä eri näkökulmista ilman, että niiden luonteen suhteen tehdään mitään ennako-oletuksia. Kohteena ovat erityisesti sellaiset tekijät, jotka yhtäältä yhdistävät valtaosaa S2-aineiston teksteistä ja toisaalta erottavat niitä S1-aineiston teksteistä (*S2-avainrakenteet* ja *Mahdollisuuden ilmaukset*), ensikieleltään erilaisten S2-kielenkäyttäjien tekstejä toisistaan (*Siirtovaikutus-avainrakenteet*) tai jotka muuttuvat vielä edistyneidenkin S2-oppijoiden teksteissä (*Pitkittäismuutokset*). Tavoitteena onkin tilastollisen avaimuuden hengessä tavoittaa tekijöitä, jotka saattavat indikoida kiinnostavia ilmiöitä – määrällisessä analyysissä paljastuvista potentiaalisista avainmuodoista ei välttämättä voi vetää suoria yhtäläisyysmerkkejä siihen kielen ilmiöön, joka todellisuudessa erottaa aineistoja toisistaan. Korpusvetoinen lähestyminen rajoittaa väistämättä sitä, missä määrin tutkimustulokset voivat ottaa huomioon lukuisat oppijankieleen ja toisen kielen oppimiseen vaikuttavat taustamuuttujat. Toisaalta itse analyysi on monilta osin korpuspohjaista: kun aineistosta on korpusvetoisesti tavoitettu erottuvia kielen muotopiirteitä, analysoidaan näiden piirteiden tyypillistä käyttöä; niiden esiintymistä erilaisissa rooleissa ja konstruktioissa sekä käytössä esiintyvää muunlaista vaihtelua. Tarvittaessa tarkastellaan myös muuta samankaltaisissa funktiossa esiintyvien kielenpiirteiden vaihtelua tai muita kielen sisäisiä tai -ulkoisia taustamuuttujia (ks. erityisesti

Pitkittäismuutokset ja Mahdollisuuden ilmaukset). Näin voidaan lopulta tavoittaa ja tulkita avainrakenteita eli sellaisia konstruktioita, joiden käyttö jotenkin erottaa edistyneitä suomenoppijoita ensikielisistä kielenkäyttäjistä tai ensikieleltään erilaisia edistyneitä suomenoppijoita toisistaan tai joiden käyttö muuttuu ajan kuluessa edistyneiden suomenoppijoiden kirjoituksessa.

4. TÄMÄN TUTKIMUKSEN AINEISTO JA MENETELMÄT

Esittelen seuraavaksi tässä väitöskirjatutkimuksessa käytetyn tutkimusaineiston sekä ne menetelmät, joita tutkimuksessa on sovellettu (käsittelen tutkimuksen kannalta relevantteja aineistollisia ja menetelmällisiä seikkoja yleisemmällä tasolla luvussa 3). Kussakin väitöskirjan osatutkimuksessa on osittain eri aineisto, minkä lisäksi korpuksen ja kehitettyjen korpustyökalujen luonne määrittävät myös käytännön tasolla tutkimusprosessia ja sitä, miten valittu tutkimusmetodi pannaan täytäntöön. Yksi väitöskirjan osatutkimuksista (*LAS2*) keskittyy yksinomaan korpuksen ja sen sisältämien työkalujen kuvaukseen, joten keskityn tässä yleisesti siihen, mitä Edistyneiden suomenoppijoiden korpus edustaa, sekä siihen, miten eri osatutkimusten aineistot keskenään erilaisten tutkimuskysymysten vuoksi eroavat toisistaan (4.1). Tämän jälkeen kuvaan vaiheittain konkreettiset menetelmälliset ratkaisuni ja kussakin vaiheessa käytetyt tilastolliset menetelmät (4.2). Aineistoa käsittelevää artikkelia lukuun ottamatta kaikki osatutkimukset (*S2-avainrakenteet*, *Siirtovaikutus-avainrakenteet*, *Pitkittäismuutokset* ja *Mahdollisuuden ilmaukset*) soveltavat avainrakenneanalyysiksi kutsumaani menetelmää jostakin näkökulmasta. Artikkelit kuitenkin poikkeavat toisistaan erilaisten tutkimuskysymysten ja sitä kautta erilaisten tutkimusasetelmien vuoksi, mistä syystä kuvaan myös tutkimusten välisiä menetelmällisiä eroja.

4.1. Edistyneiden suomenoppijoiden korpus LAS2

Kaikki tässä tutkimuksessa käytetty aineisto on osa Edistyneiden suomenoppijoiden korpus -tutkimushankkeessa koottavaa korpusta. Korpuksen kokoaminen, aineiston käsittelyyn käytettävien korpustyökalujen kehittäminen ja näiden työkalujen soveltaminen ovatkin olleet ensiarvoisen tärkeitä vaiheita tämän väitöskirjatutkimuksen mahdollistamiseksi (*LAS2*; ks. myös Ivaska & Siitonen 2009). Edistyneiden suomenoppijoiden korpus on vuonna 2007 Turun yliopistossa käynnistetty tutkimushanke, jonka tavoitteena on ollut kerätä ja saattaa tutkijayhteisön käytettäväksi suomea toisena kielenään käyttävien yliopisto-opiskelijoiden akateemisessa kontekstissa tuottamia suomenkielisiä kirjoitettuja tekstejä. Kaikista kerättävistä tekstilajeista on niin ikään kerätty tai kerätään ensikielenään suomea puhuvien yliopisto-opiskelijoiden teksteistä koostuva vertailuaineisto. Tavoitteena on ollut kerätä sellainen korpusaineisto, jonka tekstit ovat kommunikatiivisilta tavoitteiltaan autenttisia ja vertailukelpoisia ensikielisten ja toisella kielellä kirjoittavien välillä niin, että tekstejä ei ole ensisijaisesti tuotettu kielentutkimuksen aineistoksi tai osana kielitaito-opintoja. Toistaiseksi kaikki tekstit on kerätty suomen ja sen sukukielten maisteriohjelman opiskelijoilta (*S2-aineisto*) tai suomen kielen pää- ja sivuaineopiskelijoilta (*S1-aineisto*), ja

tekstit ovat tyypillisesti luonteeltaan kielitieteellisiä opintoihin liittyviin aloihin nivELYviä kulttuuri-ilmiöitä tai historiaa käsitteleviä. Korpus on luonteeltaan moniensikielinen, ja siinä on ensikielisiltä suomenpuhujilta kerätyn vertailuaineiston ohella toistaiseksi tekstejä seuraavien ensikielien puhujilta: englanti, ersä, islanti, japani, komi, liettua, puola, ruotsi, saksa, slovakki, tšekki, udmurtti, unkari, venäjä ja viro.

Korpus on jaettu kolmeen osakorpukseen sen mukaan, mikä on tekstien ensisijainen funktionaalinen pyrkimys. Osakorpukset ovat: 1) tenttivastaukset; 2) opinnäytteiksi tai julkaistaviksi tarkoitetut tekstit ja 3) opiskeluun ja oppimiseen nivELYvät tekstit, kuten kurssiesseet ja oppimispäiväkirjat. Osakorpukset poikkeavat toisistaan niin tekstien oletetun lukijakunnan, niiden ensisijaisen funktion, tuottamistilanteen kuin konkreettisten kirjoitusvälineidenkin osalta. Olen käyttänyt tämän väitöskirjan kaikissa osatutkimuksissa tenttivastaus-osakorpusta ja kuvaan seuraavaksi sen luonnetta. Tenttivastaukset on kirjoitettu valvotuissa olosuhteissa, ja ne on kirjoitettu kynällä paperille. Tekstien ensisijaisena kommunikatiivisena tavoitteena on osoittaa tentin vastaanottajalle kirjoittajan tiedot tentittävästä sisällöstä – joko kuunneltuihin luentoihin tai luettuun kirjallisuuteen perustuen. Suomen ja sen sukukielten maisteriohjelman opettajien ja opiskelijoiden välillä on lisäksi erikseen sovittu, että tenttivastauksen kielellinen ilme ei vaikuta tentin arvosteluun niin kauan, kuin tentin vastaanottaja kokee ymmärtävänsä, mitä kirjoittaja tarkoittaa. Tenttien tehtävänannot on tallennettu aineistoon silloin, kun ne ovat olleet saatavilla, mutta yksittäisten tehtävänantojen vaikutusta tutkittaviin teksteihin ei tarkastella tämän tutkimuksen puitteissa.

Aineistona käytetty materiaali on annotoitu kattavasti: se on rakenteistettu kappaleisiin, virkkeisiin, lauseisiin ja sanoihin, ja sanatason elementit on annotoitu sanojen perusmuotojen, sanaluokkien, morfologisten muotojen ja syntaktisten funktioiden osalta. Lisäksi teksteistä on erotettu mahdolliset tenttikysymysten, välimerkkien sekä kielellisten esimerkkien kaltaiset jaksot siten, että korpuksella tehtävät haut eivät kohdistu niihin mutta ne ovat nähtävissä aineistoa tarkasteltaessa. Korpus on tallennettu XML-muotoon, joka seuraa muissa Turun yliopistossa kootuissa korpuksissa sovellettua mallia (ks. LAX). Asetelma 1 kuvaa korpuksen sanatason annotaation eri kerroksia.

Virke:	Liikkuva	väki	tarvitsi	nimeä.
Lemma:	LIKKUA	VÄKI	TARVITA	NIMI
Sanaluokka:	V	N	V	N
Morfologia:	pcp1 sg nom	sg nom	fin ind pret sg3	sg part
Syntaksi:	substantiivin määrite	subjekti	predikaatti	objekti

Asetelma 1. Virkkeen *Liikkuva väki tarvitsi nimeä.* sanatason annotaatio Edistyneiden suomenoppijoiden korpuksessa.

Suomen ja sen sukukielten maisteriohjelman kieli on suomi, ja opiskelijoilta vaaditaan pääsääntöisesti kandidaattitason tutkintoa vastaavia opintoja Suomen kielessä ja kulttuurissa. Opiskelijoiden odotetaan kykenevän tekemään kaikki tarvittavat akateemiset suoritukset suomen kielellä. Opiskelijoilta ei kuitenkaan edellytetä erillistä kielitaitotestiä, vaan kielitaidon osalta arvio perustuu edeltäviin opintoihin ja hakudokumentteihin. Kun näitä odotuksia vertaa Eurooppalaisen viitekehyksen kielitaitotasoihin (CEF) perustuvan CEFLING-tutkimushankkeen funktionaalisiin määritelmiin (ks. esim. Kajander 2013: 238–240), voidaan informanttien kirjoitustaidon olettaa olevan vähintään B2-tasolla – kirjoittajat pystyvät tehokkaasti ilmaisemaan tietoja ja näkemyksiä ja suhteuttamaan niitä muualta saatuun tietoon ja muiden näkemyksiin. Tenttivastaus-osakorpuksen osalta jokaiselta informantilta on lisäksi arvioitu viitekehyksen asteikolle vähintään kaksi tekstiä, ja arviot vahvistavat tämän oletuksen (tähän mennessä arvioitujen tekstien jakautuminen eri taitotasolle: B1: 4 %, B2: 32 %, C1: 62 %, C2: 3 %).

Eri osatutkimuksissa käytetyt aineistot eroavat toisistaan joiltakin osin – pääosin siksi, että kysymyksenasettelu on eri osatutkimuksissa erilainen. Kun artikkeleissa *S2-avainrakenteet* ja *Mahdollisuuden ilmaukset* keskeinen vertailu käydään ensikielisyys- ja ei-ensikielisyys välillä, artikkelissa *Siirtovaikutus-avainrakenteet* vertaillaan japania, liettuaa, tšekkiä, uncaria ja venäjää ensikielisinä puhuvien informanttien tekstejä toisiinsa ja artikkelissa *Pitkittäismuutokset* vertaillaan tekstejä suhteessa siihen, kuinka monta kuukautta saman kirjoittajan ensimmäisen korpuksen tallennetun tekstin jälkeen kulloinenkin teksti on kirjoitettu. Nämä muuttujat puolestaan rajaavat osan aineistosta pois, sillä esimerkiksi kaikkien tekstien tarkkaa kirjoitushetkeä ei tiedetä (ks. tarkemmat kriteerit kustakin artikkelista). Taulukko 1 kuvaa kunkin osatutkimuksen aineiston kokoa.

Taulukko 1. Väitöskirjan osatutkimusten aineistot. S1-aineiston kaikki tekstit ovat eri kirjoittajilta.

Osatutkimus	S2-aineiston saneiden määrä	S2-aineiston tekstien määrä	S2-aineiston informanttien määrä	S1-aineiston saneiden määrä	S1-aineiston tekstien määrä
<i>S2-avainrakenteet</i>	120 965	275	31	30 399	56
<i>Siirtovaikutus-avainrakenteet</i>	japani: 8 498 liettua: 11 397 tšekki: 8 745 unkari: 10 885 venäjä: 9 721 (yht: 49 246)	japani: 20 liettua: 20 tšekki: 20 unkari: 20 venäjä: 20	japani: 4 liettua: 4 tšekki: 4 unkari: 4 venäjä: 4	10 283	20
<i>Pitkittäismuutokset</i>	86 372	196	37	58 333	102
<i>Mahdollisuuden ilmaukset</i>	145 379	334	26	58 649	102

4.2. Avainrakenneanalyysi

Kielenpiirteiden määrittely perustuu tämän tutkimuksen teoreettisen viitekehyksen mukaisesti konstruktiokieliopin ajatukseen siitä, että abstraktiotasoltaan ja laajuudeltaan erilaiset kielenpiirteet yksittäisistä foneemeista, morfeemeista ja lekseemeistä argumenttirakenteen kaltaisiin abstrakteihin lausekokonaisuuksiin eivät kieliopin kannalta eroa toisistaan perustavanlaatuisesti (ks. 2.1.2). Kutsun soveltamaani menetelmää avainrakenneanalyysiksi (ks. osatutkimusten ohella myös Ivaska & Siitonen 2011). Sovellan pääpiirteissään kontrastiivisen välikieliansalyysin ajatusta tiettyjen muuttujien mukaan eroteltujen samankielisten aineistojen määrällisestä vertailusta. Menetelmä pyrkii kuitenkin tarjoamaan konkreettisia keinoja havaittujen erojen taustalla vaikuttavien kielellisten ja kielenulkoisten tekijöiden analyysiin. Prosessi on monivaiheinen ja -menetelmäinen, ja siinä yhdistyy piirteitä niin rinnakkaisesta (engl. *concurrent*) kuin peräkkäisestäkin (engl. *sequential*) monimenetelmäisyydestä (ks. käsitteistä esim. Hashemi 2012). Menetelmän perusidea on siis seuraava:

- 1) paikannetaan tilastollisen tarkastelun avulla vertailtavien aineistojen väliltä suurimmat ja vertailtavia aineistoja parhaiten erottavat kielenpiirteiden frekvenssierot;
- 2) selvitetään tavoitettujen piirteiden sisäisen vaihtelun ja niiden tyypillisen esiintymisympäristön analyysin avulla, millaisissa kielellisissä rakenteissa kyseiset kielenpiirteet tyypillisesti esiintyvät;
- 3) selvitetään löydettyjen rakenteiden ja tarvittaessa muiden analyysissä esiin nousseiden relevanttien rakenteiden tyypillistä käyttöä ja käytön eroja sekä eroja määrittäviä kielellisiä ja kielenulkoisia tekijöitä eri osa-aineistoissa.

Käytän tutkimuksen lähtökohtana aineiston morfologista annotaatiota ja käytetyissä morfologisissa muodoissa havaittavia frekvenssieroja tilastollisten testien avulla havaittavien avainmuotojen tavoittamiseksi. Aineiston kustakin tekstistä siis lasketaan korpusta varten kehitettyjen hakutyökalujen avulla (ks. LAS2) erikseen niissä esiintyvien morfologisten 1-, 2-

ja 3-skipgrammien (ks. 2.1.4) frekvenssit, jotka normalisoidaan kuvaamaan grammien frekvenssiä tuhatta sanaa kohti. Näin jokaiselle aineistossa esiintyvälle grammille saadaan sen suhteellinen frekvenssi jokaisessa aineiston tekstissä. Olen valinnut morfologisen annotaation tutkimuksen lähtökohdaksi. Näin lähtökohtana olevat yksiköt ovat funktionaalisten tulkintojen (kuten objekti) sijaan ainakin osin myös konkreettisen muotonsa perusteella määriteltyjä kielenaineiksia (kuten objektin tunnuksena käytetty partitiivi). Lisäksi tutkimus ottaa näin paremmin huomioon suomen taiputukseltaan rikkaan rakenteellisen luonteen.

Tutkimusasetelma on vertaileva, ja siinä sovelletaan tilastollista avaimuutta avainmuotojen paikallistamiseen. Toisin sanoen grammit järjestetään tilastollisten menetelmien avulla laskevaan järjestykseen sen mukaan, mitkä ovat kunkin osatutkimuksen kannalta vahvimpia potentiaalisia avainmuotoja eli minkä muotojen frekvenssissä on yleisesti suurimmat erot. Käytettynä tilastollisena menetelmänä on niin sanottu satunnaismetsä-menetelmä (engl. *random forest*), joka on alun perin suunniteltu datan automaattiseen kategoriointiin (algoritmista tarkemmin ks. Breiman 2001). Menetelmän avulla voidaan etsiä muuttujia, jotka ennustavat parhaiten aineiston eri alaryhmiä tai joiden arvot korreloivat jonkin toisen muuttujan kanssa (menetelmän soveltamisesta kielitieteelliseen tutkimukseen ks. esim. Tagliamonte & Baayen 2012). Satunnaismetsä-menetelmän etuna pidetään sen joustavuutta: se ei aseta ehtoja aineiston jakaumille eikä tarkasteltavien muuttujien luonteelle. Kahden piirteen välisten frekvenssien vertailuissa sovelletaan tarvittaessa lisäksi Mann-Whitneyn U-testiä, joka on luonteeltaan epäparametrinen, mistä syystä aineiston jakauman ei tarvitse olla symmetrinen. Osatutkimuksessa *Siirtovaikutus-avainrakenteet* sovelletaan lisäksi Tukeyn testiä eri ensikieliryhmien välisten frekvenssierojen luonteen selvittämiseksi.

Suurimpien erojen paikantamisen jälkeen aineistosta haetaan näiden muotojen kaikki esiintymät ja tarkastellaan esiintymien sisäistä vaihtelua – mitkä lekseemit ja minkä sanaluokkien edustajat esiintyvät kyseisessä avainmuodossa ja missä syntaktisissa funktioissa muoto esiintyy – sekä niiden tyypillistä esiintymisympäristöä – mitä lekseemejä ja morfologisia muotoja, minkä sanaluokkien edustajia ja mitä syntaktisia funktioita esiintyy havaitun potentiaalisen avainmuodon kielellisessä lähikontekstissa (*S2-avainrakenteet*, *Siirtovaikutus-avainrakenteet* ja *Pitkittäismuutokset*). Käytännössä muotoja tarkastellaan siis kollostruktionaalisen, kollokationaalisen ja kolligationaalisen analyysin avulla (ks. 2.1.4). Havaittujen tendenssien ja niissä nähtyjen erojen avulla voidaan päätellä, mitkä ovat ne konstruktiot, joissa havaitut avainmuodot esiintyvät sekä mikä tai mitkä näistä

konstruktiosta todennäköisimmin ovat havaitun eron taustalla ja ovat siis avainrakenteita. Eroja selittävien tekijöiden tilastolliseen havainnointiin käytetään satunnaismetsämenetelmää ja havaittujen erojen tilastollisen merkitsevyyden testaamiseen Mann-Whitneyn U-testiä. Lisäksi kollostruktioiden tilastollisissa analyyseissa (*Pitkittäismuutokset*) sovelletaan Fisherin tarkkaa testiä.

Potentiaalisiksi avainmuodoksi havaitun morfologisen grammin analysoinnin ohella analysoidaan laadullisesti sellaisten konstruktioiden käyttöä, jotka vaikuttavat joko määrällisten tai laadullisten havaintojen perusteella todennäköisiksi eron taustalla vaikuttaviksi rakenteiksi. Havaitun konstruktion lisäksi huomioon otetaan tarvittaessa siihen funktionaalisesti tai semanttisesti mahdollisesti liittyvät muut konstruktiot sekä näiden konstruktioiden ja tekstien suhde niiden edustamaan tekstilajiin sekä kielenulkoiseen todellisuuteen. Näin voidaan tarkastella myös niitä syitä, jotka ovat havaittujen erojen takana. Näitä syitä käsitellään erityisesti osatutkimuksissa *Pitkittäismuutokset* ja *Mahdollisuuden ilmaukset*, joskin mahdollisia syitä nostetaan esiin myös muissa osatutkimuksissa.

Väitöskirjan eri osatutkimukset keskittyvät osin avainrakenneanalyysin eri vaiheisiin, mistä syystä myös menetelmän käytännön sovellukset poikkeavat hieman toisistaan. Siinä missä *S2-avainrakenteet* nostaa esiin niitä avainmuotoja, jotka erottavat S2- ja S1-aineistoja toisistaan, ja selvittää niitä konstruktiota, jotka ovat eron taustalla, *Mahdollisuuden ilmaukset* paneutuu yhteen näistä havaituista konstruktiosta ja tarkastelee tarkemmin sitä ja sen suhdetta semanttisesti lähimerkityksisiin muihin konstruktiioihin, niiden kielenulkoiseen käyttöympäristöön ja yhdistää eron syitä käyttöpohjaisen mallin perusoletuksiin. *Pitkittäismuutokset* puolestaan kuvaa koko menetelmällisen prosessin mutta keskittyy analyysissä ainoastaan yhteen havaittuun potentiaaliseen avainrakenteeseen. Olen lisäksi tarvittaessa muokannut asetelmaa niin, että se olisi mahdollisimman yhteismitallinen aiemman menetelmällisesti samankaltaisen tutkimuksen kanssa (ks. erityisesti *Siirtovaikutus-avainrakenteet*). Lähtökohtanani on ollut tavoittaa aineistosta korpusvetoisesti sellaiset sanojen morfologiset muodot tai morfologisten muotojen yhdistelmät, joita käytetään S2-aineistossa selvästi S1-aineistoa vähemmän tai enemmän (*S2-avainrakenteet*), joiden käyttöfrekvenssi erottaa ensikielitaustaltaan erilaisia S2-oppijoita toisistaan (*Siirtovaikutus-avainrakenteet*) tai joiden suhteellinen frekvenssi muuttuu tarkastelujakson aikana (*Pitkittäismuutokset*).

5. TULOKSET

Kuvaan tässä luvussa väitöskirjatutkimukseni keskeiset tulokset. Väitöskirjan osatutkimukset limittyvät tulosten kannalta toisiinsa monissa kohdin, joten en esittele tuloksia yksinomaan osatutkimuksittain. Luon ensin (5.1) määrällisen yleiskuvan niihin tilastollisesti tavoitettuihin morfologisiin avainmuotoihin, jotka käyttöfrekvenssinsä perusteella erottavat 1) edistynyttä oppijansuomea ensikielisten kirjoittajien kielestä (*S2-avainrakenteet*) tai 2) eri ensikieliä puhuvia S2-oppijoita toisistaan (*Siirtovaikutus-avainrakenteet*), sekä niihin avainmuotoihin, 3) joiden frekvenssi muuttuu eniten pitkittäistarkastelussa (*Pitkittäismuutokset*). Tämän jälkeen esittelen tarkemmin löydettyjä avainrakenteita eli niitä konstruktioita, joiden käyttö havaittujen avainmuotojen analyysin perusteella erottaa edistyneitä suomenoppijoita ensikielistä kirjoittajista (5.2; *S2-avainrakenteet* ja *Pitkittäismuutokset*), jotka erottavat eri ensikieliä puhuvia suomenoppijoita (5.3; *Siirtovaikutus-avainrakenteet*) tai joiden käyttö muuttuu ajan kuluessa (5.2; *Pitkittäismuutokset*). Näin tulokset valottavat käytetyn korpusvetoisen menetelmän soveltuvuutta oppijankielen tutkimukseen kolmelta alalle tyypillisen kysymyksenasettelun kannalta. Lopuksi tarkastelen kahden tapaustutkimuksen kautta niitä syitä, jotka mahdollisesti vaikuttavat havaittujen erojen taustalla sekä näiden syiden käyttöpohjaista luonnetta (5.4; *Pitkittäismuutokset* ja *Mahdollisuuden ilmaukset*).

5.1. Määrällinen yleiskuva havaituista potentiaalisista avainmuodoista

Etsiessäni S2-aineistoa ja S1-aineistoa toisistaan erottavia morfologisia avainmuotoja osatutkimuksessa *S2-avainrakenteet* paneuduin kaikista aineistossa esiintyneistä morfologisista 1-, 2- ja 3-grammeista 50:een esiintymisfrekvenssin perusteella osa-aineistoja toisistaan parhaiten erottavaan grammiin. Rajasin lisäksi näistä 50 grammista käsittelyn ulkopuolelle ne 2- ja 3-grammit, joiden jokin osa tai jotkin osat ovat jo itsessään edustettuina 50:n parhaiten erottavan grammin joukossa – nämä grammit edustavat luultavimmin erojen tyypillisiä esiintymisympäristöjä. Näillä perustein aineistosta erottuu potentiaalisina S1- ja S2-aineistoa toisistaan erottavina avainmuotoina yhteensä viisi 1-grammia (vA-partisiipin yksikön genetiivi <pcp1 sg gen> (esim. *ajattelevan*), finiittimuotoisen indikatiivissa olevan verbin preteritin yksikön kolmas persoona <fin ind pret sg3> (esim. *tuli*), finiittimuotoisen konditionaalissa olevan verbin preesensin yksikön kolmas persoona <fin cond pres sg3> (esim. *tulisi*), yksikön translatiivi <sg tra> (esim. *esimerkiksi*) sekä A-infinitiivi <inf1> (esim. *tulla*)), kaksi 2-grammia (finiittimuotoisen indikatiiviverbin preesensin yksikön kolmannen persoonan ja yksikön genetiivin muodostama <fin ind pres sg3><sg gen> (esim. *kantaa maailman*) sekä finiittimuotoisen passiivisen indikatiiviverbin preesensin ja yksikön illatiivin

muodostama <fin pass ind pres><sg ill> (esim. *etsitään hypoteesiin*) ja yksi 3-grammi (monikon partitiivista, finiittimuotoisen indikatiiviverbin preesensin monikon kolmannesta persoonasta ja monikon nominatiivista koostuva <pl part><fin ind pres pl3><pl nom> (esim. *indeksisiä ovat erilaiset*)). Taulukko 2 kuvaa näiden muotojen tekstikohtaisten frekvenssien keskiarvoja S1- ja S2-aineistoissa. Avainmuodoista seitsemän on frekventimpiä S1-aineistossa, eli ne ovat S2-aineistossa aliedustuneita S1-aineistoon nähden, kun taas yksi avainmuoto (<fin ind pret sg3>) on frekventimpi eli yliedustunut S2-aineistossa S1-aineistoon nähden.

Taulukko 2. S1- ja S2-aineistoja toisistaan parhaiten erottavat n-grammit ja niiden frekvenssien keskiarvot. Grammit ovat taulukossa piteuden mukaan järjestettynä 1-grammeista alkaen, ja samanmittaiset grammit ovat siinä järjestyksessä, miten hyvin ne erottavat aineistoja toisistaan.

n-grammi	Keskiarvo / 1000 sanetta	
	S1	S2
<pcp1 sg gen>	2,9	0,7
<fin ind pret sg3>	4,5	21,5
<fin cond pres sg3>	5,6	1,9
<sg tra>	12,0	6,5
<inf1>	37,4	22,9
<fin ind pres sg3><sg gen>	35,8	22,1
<fin pass ind pres><sg ill>	2,1	0,8
<pl part><fin ind pres pl3><pl nom>	1,7	0

Osatutkimuksessa *Siirtovaikutus-avainrakenteet* etsin morfologisia avainmuotoja, jotka erottavat eri ensikieliä puhuvia edistyneitä suomenoppijoita toisistaan. Olen rajannut tarkastelun niihin 1-, 2- ja 3-grammeihin, jotka ovat 30:n osa-aineistoja parhaiten toisistaan erottavan grammin joukossa. Olen lisäksi rajannut käsittelyn ulkopuolelle ne 2- ja 3-grammit, joiden jokin osa sisältyy johonkin muuhun näistä 30 grammista, ne, joissa yksittäisten kirjoittajien väliset erot selittävät eroa paremmin kuin ensikielten väliset erot, sekä ne, joiden frekvenssissä ei valikoitumisesta huolimatta havaita tilastollisesti merkitsevää eroa minkään ensikieliryhmän välillä. Näillä perustein avainmuodoiksi erottuu kaksi 1-grammia (konjunktio <cnj> (esim. *ja*) sekä finiittimuotoisen indikatiiviverbin preesensin yksikön kolmas persoona <fin ind pres sg3> (esim. *tulee*) ja yksi 2-grammi (yksikön nominatiivista ja monikon partitiivista koostuva <sg nom><pl part> (esim. *hän muita*)). Taulukko 3 kuvaa näiden muotojen tekstikohtaisten frekvenssien keskiarvoja tarkastelluissa osa-aineistoissa. Ensikieliryhmät ryhmittyvät avainmuotojen frekvenssin kannalta eri grammeissa eri tavoin. Konjunktion osalta japania ja unkaria ensikielenään puhuvien osa-aineistot ovat lähellä toisiaan ja vertailuaineistoa – niissä käytetään konjunktioita selvästi tšekkiä, liettuaa ja venäjää puhuvien osa-aineistoja enemmän. Indikatiivissa olevan verbin kohdalla sen sijaan

unkaria ensikielenä puhuvien osa-aineisto ja vertailuaineisto käyttäytyvät keskenään samoin ja eroavat etenkin liettuaa ensikielenä puhuvien osa-aineistosta. Yksikön nominatiivista ja monikon partitiivista koostuvan grammin käytössä puolestaan japania, tšekkiä, unkaria ja liettuaa ensikielinen puhuvien osa-aineistot ovat lähellä toisiaan ja eroavat jossain määrin venäjää ensikielenä puhuvien aineistosta ja erityisesti vertailuaineistosta.

Taulukko 3. Eri ensikielisten kirjoittajien tekstejä toisistaan eniten erottavat n-grammit ja niiden frekvenssien keskiarvot. Vertailuaineiston (fi) lisäksi tarkastellut ensikielet ovat tšekki (cs), unkari (hu), japani (ja), liettua (lt) ja venäjä (ru).

n-grammi	Keskiarvo / 1000 sanetta					
	S1-fi	S2-cs	S2-hu	S2-ja	S2-lt	S2-ru
<cnj>	95,6	76,5	94,5	97,3	74,4	76,5
<fin ind pres sg3>	93,8	77,3	94,9	82,0	64,2	77,3
<sg nom><pl part>	24,7	9,8	9,8	9,4	11,9	17,1

Pitkittäismuutokset-osatutkimuksessa paneuduin ainoastaan siihen avainmuotoon, jonka frekvenssin muutokset korreloivat S2-aineistossa selvimmin kuluvaan ajan kanssa kuudentoista kuukauden tarkastelujakson aikana. Taulukko 4 kuvaa keskimääräistä frekvenssiä tarkastelujakson eri vaiheissa sekä sen frekvenssiä vertailuaineistossa. Täksi muodoksi erottuu finiittimuotoisen indikatiiviverbin preteritin yksikön kolmas persoona (<fin ind pret sg3> (esim. *tuli*)). Muodon keskimääräinen frekvenssi lähes puolittuu tarkastelujakson aikana, ja se lähenee samanaikaisesti muodon keskimääräistä frekvenssiä S1-aineistossa. Muoto on siis sama, joka erottui avainmuodoksi myös S1- ja S2-aineiston välisessä vertailussa. Muodon frekvenssi eri osatutkimusten S1-aineistoissa vaihtelee hieman eri aineisto-otannan vuoksi (vrt. taulukot 2 ja 4), mutta se on kuitenkin aivan eri luokkaa kuin S1- ja S2-aineistojen välinen ero.

Taulukko 4. Finiittimuotoisen indikatiiviverbin preteritin yksikön kolmannen persoonan (<fin ind pret sg3>) frekvenssin keskiarvot 16 kuukauden tarkastelujakson eri vaiheissa ja vertailuaineistossa.

n-grammi	Keskiarvo / 1000 sanetta			
	0-4 kk	5-10 kk	11-15 kk	S1-aineisto
<fin ind pret sg3>	23,3	21,6	10,6	3,2

5.2. Edistynyttä oppijansuomea ensikielisestä suomesta erottavat konstruktio

Edistynyttä S2-suomea ja ensikielistä suomea erottavat konstruktio voidaan tämän aineiston osalta jakaa neljään ryhmään sen mukaan, minkälaisia havaitut erot ovat luonteeltaan. On kuitenkin selvää, että nämä lankeavat osin myös yhteen. Osa eroista näyttää liittyvän

ensikielisen aineiston yleisesti suurempaan morfosyntaktiseen kompleksisuuteen, osa modaalisuuden runsaampaan ilmaisemiseen ensikielisessä aineistossa (ks. myös 5.4.1), osa tiettyihin leksikaalisesti rajattuihin konstruktioihin ja osa aikamuotojen käyttöeroihin aineistoissa (ks. myös 5.4.2). Havaitut erot ovat luonteeltaan, skemaattisuudeltaan ja merkitykseltään hyvin erilaisia konstruktioita, mutta ne voidaan kaikki nähdä muodon ja funktion toistuvina vakiintuneina yhdistelminä, mikä puolestaan puoltaa tämän väitöskirjatutkimuksen konstruktioilähestymistä. Monet havaituista konstruktioista ovat luonteeltaan sellaisia, että niiden käyttöero saattaa ainakin joiltakin osin liittyä käsillä olevan tekstilajin tyypillisiin kielellisiin konventioihin ja niiden konventioiden oppimiseen ja sitä kautta kielen, kieliopin ja kielitaidon kontekstisidonnaisuuteen ja käyttöperustaisuuteen (ks. 5.4). Havaituista eroista valtaosa on sellaisia, että jokin kielenpiirre on S2-aineistossa S1-aineistoa harvinaisempi. Monet havaituista piirteistä ovat verraten harvaan esiintyviä jo itsessään, ja niiden korkeampi frekvenssi S1-aineistossa tukeekin ylipäätään ajatusta S1-kielenkäyttäjien suuremmasta kompleksisuudesta ja vaihtelusta.

Indikatiivisesta finiittiverbistä ja yksikön genetiivistä muodostuvan 2-grammin käyttöero vaikuttaa yksikön ja sen ympäristön analyysin perusteella nivoutuvan yhteen sen kanssa, että S2-aineistossa käytetään S1-aineistoa vähemmän substantiivin määritteitä erityisesti verbinjälkeisessä asemassa. Esimerkki 18 kuvaa yhtä tällaista lausetta.

18) *Se on sanan perusosa ilman taivutuspäätteitä. (S2: las2-7tt01te027⁵)*

Finiittimuotoisten yksikön kolmannen persoonan preesensissä olevien verbien sen kummemmin kuin genetiivimuotoisten substantiivinmääritteiden määrässäkään ei ole mainittavaa eroa aineistojen välillä, mutta S2-aineistossa on suhteellisesti S1-aineistoa vähemmän valinnaisia nominaalisia jäseniä verbinjälkeisessä asemassa, kun taas verbinjälkeiset substantiivinmääritteet sijaitsevat S2-aineistossa suhteellisesti S1-aineistoa useammin lähellä lauseen loppua. Havainto voidaan tulkita S1-aineiston suuremmaksi morfosyntaktiseksi kompleksisuudeksi ja vaihteluksi. Myös aiempi samaa korpusta käyttänyt tutkimus tukee tulkintaa: edistyneessä oppijansuomessa on osoitettu olevan ensikielisiä tekstejä vähemmän lausekkeita, joissa on vähintään kaksi määritettä (Kannisto 2012), vaikka lauseiden kokonaispituudet eivät eroa toisistaan merkittävästi (Salmi 2010: 62–63), minkä

⁵ Merkinnän alkuosa (tässä S2) kertoo sen osa-aineiston, johon esimerkki kuuluu ja loppuosa (tässä las2-7tt01te027) kertoo sen sijainnin korpuksessa. Käytän samaa merkintätapaa kaikissa aineistoesimerkeissäni.

lisäksi substantiivilausekkeiden määritteiden on ylipäättään osoitettu lisääntyvän kielitaidon kehittyessä (Ukkola 2009). Näiden tutkimusten aineisto on nyt käsillä olevaa tutkimusta pienempi, mutta ainakin pääosa niiden aineistosta sisältyy myös tämän tutkimuksen aineistoon, joten tuloksia voinee pitää vertailukelpoisina.

VA-partisiipin yksikön genetiivin (<pcp1 sg gen>) suuremman määrän S1-aineistossa voi niin ikään tulkita liittyvän aineiston suurempaan morfosyntaktiseen kompleksisuuteen. Yksikön ja sen ympäristön analyysi osoittaa, että S2-aineistossa huomattavasti pienempi osa muodon esiintymistä on esimerkin 19 kaltaisia referatiivikonstruktioita. Muoto on verrattain harvinainen kummassakin aineistossa, mutta havaittu frekvenssiero selittyy nimenomaan tämän konstruktion vähäisellä käytöllä S2-aineistossa.

19) Suomen kielen taito on osoitettu *heikkentyvän* (S2: las2-30tt01te06)

Kolmanneksi morfosyntaktiseen kompleksisuuteen liittyväksi eroksi tulkitsen esimerkin 20 kaltaisten, monikon partitiivista, monikollisesta indikatiivissa olevasta verbistä ja monikon nominatiivista (pl part><fin ind pres pl3><pl nom>) koostuvien, sanajärjestykseltään käänteisten predikatiivikonstruktioiden runsaamman käytön S1-aineistossa.

20) Yksi tapa on luokitella ne niiden paikkaisuuden perusteella. *Nollapaikkaisia ovat verbit [...]* (S1: las2-vtt01vert001)

Suoran sanajärjestyksen predikatiivilauseissa sen kummemmin kuin suoran tai käänteisen sanajärjestyksen transitiviilauseissa ei ole huomattavia eroja aineistojen välillä. Esimerkin 20 kaltaiset inversiot liittyvät tekstin sidosteisuuden luomiseen, mitä ilmaisevat keinot puolestaan saattavat aiemman tutkimuksen mukaan erottaa kielitaitotasoltaan erilaisia kirjoittajia toisistaan (ks. Komppa 2012: 186–187). Osin ero liittyy myös monikollisten predikatiivien sijamerkintään sekä siihen mahdollisesti liittyvään subjektin ja predikatiivin välisten roolien hämärtymiseen, sillä S2-aineistossa yli puolet sanajärjestykseltään käänteisten monikollisten predikatiivikonstruktioiden predikatiiveista on nominatiivissa kun taas S1-aineistossa kaikki tapaukset ovat partitiivissa. Tämä puolestaan vastaa aiemmassa tutkimuksessa osoitettua tendenssiä siitä, että useampien prototyypisistä konstruktiosta poikkeavien piirteiden lankeaminen yhteen lisää muotopiirteiltään toisiaan lähellä olevien konstruktioiden sekoittumisen todennäköisyyttä oppijansuomessa (ks. Ivaska 2011).

Modaalisuuden ilmaiseminen näyttäytyy niin ikään aineistoja erottavana piirteenä, sillä sekä esimerkkien 21 ja 22 kaltaiset *voida*-verbistä ja A-infinitiivistä (ks. myös 4.5) että esimerkin 23 kaltaiset genetiivisestä subjektista ja *tulla*-verbin konditionaalimuodosta muodostuvat konstruktiot ovat S1-aineistossa S2-aineistoa yleisempiä. Aiempi tutkimus tukee tulkintaa, sillä niin passiivissa olevien modaaliverbien kuin *voi*-konstruktioidenkin on todettu yleistyvän kielitaidon karttuessa (Seilonen 2013), minkä lisäksi modaalisten verbiketjujen käytön on todettu erottavan kaksikielisessä ympäristössä ja luokkahuoneympäristössä suomea oppineita kielenoppijoita toisistaan (Niiranen 2008).

21) Opettaja *voi puhua* helpotettua kieltä [...] (S1: las2-vtt01vert050)

22) Ennen kaikkea sanat *voidaan jakaa* niiden taivutusmuotojen perusteella. (S2: las2-14tt01te05)

23) Kielen *oppimisen tulisi* tähdätä käytännönläheisissä kommunikaatiotilanteissa pärjäämiseen (S1: las2-vtt01vert058)

Tutkimuksessa erottui myös kolme sellaista leksikaalisesti rajattua konstruktiota, joita käytetään S1-aineistossa S2-aineistoa enemmän. Yksikön translatiivin frekvensseissä havaittu ero selittyy yksinomaan esimerkin 24 kuvastaman *esimerkiksi*-sanamuodon runsaammalla käytöllä S1-aineistossa, kun taas esimerkit 25 ja 26 kuvaavat kahta eri konstruktiota, jotka molemmat koostuvat passiivisesta verbistä ja illatiivimuotoisesta substantiivista. Yksikön ja sen ympäristön analyysi osoittaa, että suuri osa näiden muotojen esiintymistä on joko esimerkin 25 kaltaisissa transitiivisissa luokittelukonstruktioissa, jotka toteuttavat karkeasti skeemaa 'X jaetaan/luokitellaan/järjestetään/jne. Y:ihin / N:ään Y:hyn / Y:hyn ja Z:aan', tai esimerkin 26 kaltaisissa prosessia kuvaavissa intransitiivisissa konstruktioissa, jotka toteuttavat karkeasti skeemaa 'X:ssä viitataan/keskitytään/pureudutaan/jne. Y:hyn'.

24) Arvioimista voi vaikeuttaa *esimerkiksi* se, [...] (S1: las2-vtt01vert065)

25) Lausesemantiikkaa tarkasteltaessa sanat *jaetaan* sisältö- ja *muotosanoihin* (S1: las2-vtt01vert025)

26) *Demonstratiivideiksiksessä* viitataan kieliaktissa havaittavaan tai tekstissä ja kontekstissä mainittuun *tarkoitteeseen* (S2: las2-6tt01te08)

Niin ikään aikamuotojen käyttö erottaa S1-aineistoa ja S2-aineistoa, joskin aikamuotojen välinen distribuutio muuttuu selvästi pitkittäistarkastelussa. Esimerkin 27 kaltaiset preteritiverbit ovat S2-aineistossa selvästi S1-aineistoa yleisempiä, kun taas preesensissä olevia verbejä on S1-aineistossa selvästi S2-aineistoa enemmän. Muutoksia tapahtuu ennen kaikkea eri aikamuotojen välisessä distribuutiossa, sillä preteritiä käytetään kuudentoista kuukauden tarkastelujakson ensimmäisenä viitenä kuukautena keskimäärin yli kaksi kertaa

niin paljon (34,5 / 1000 sanetta) kuin jakson viimeisenä viitenä kuukautena (15,7 / 1000 sanetta) verrattuna. Tilastollisen analyysin perusteella tekstien aihepiirit tai informanttien välinen vaihtelu eivät selitä eroa. Samanaikaisesti preesensin käyttö yleistyy vastaavalla tavalla (ensimmäiset viisi kuukautta: 100,8 / 1000 sanetta; viimeiset viisi kuukautta: 117,1 / 1000 sanetta), ja käyttö lähenee määrällisesti S1-aineistoa (preteriti: 5,5 / 1000 sanetta; preesens: 120,3 / 1000 sanetta).

27) Gallen [...] *kuvatti Kalevalaa (las2-23tt01te19)

Yksikön ja sen ympäristön analyysin perusteella ero perustuu nimenomaan aikamuotojen väliseen distribuutioeroon, ei mihinkään yksittäiseen semanttiseen ryhmään, minkään yksittäisten tai harvojen lekseemien systemaattiseen käyttöeroon tai mihinkään rajattujen konstruktoiden käyttöeroon tietyissä aikamuodoissa. Käyttö ei tarkastelujakson aikana kuitenkaan lähene ainakaan kaikilta osin laadullisesti S1-aineistossa havaittua tyyppillistä käyttöä. Tältä osin tulokset eivät tämän aineiston osalta tue aiemmassa toisen kielen tutkimuksessa havaittua verbien leksikaalisen aspektin vaikutusta aikamuotoihin ja niiden oppimiseen (vrt. esim. Bardovi-Harlig 2000; Wulff ym. 2009), joskaan leksikaalisen aspektin vaikutusta ei voida myöskään sulkea pois. DEMfad-mallin näkökulmasta voidaan sanoa, että tarkastelunalaiset konstruktioit sinänsä ovat informanteille tuttuja jo ennen tarkastelujakson alkua. Niiden tyyppilliset käyttöfunktioit ja sitä kautta käyttöfrekvenssit eroavat kuitenkin ensikielisestä vertailuaineistosta, eikä käytön tältä osin voi sanoa olevan tarkkuudeltaan kohdekielen kaltaista. Käytön määrällisen ja funktionaalisen kohdekielistymisen voitaneen puolestaan katsoa tarkoittavan tekstilajille tyyppillisten kielellisten konventioiden oppimista.

5.3. Ensikielikohtaiset erot edistyneen oppijansuomen konstruktioissa

Eri ensikieliryhmien välisiä eroja käsittelevässä osatutkimuksessa nousi esiin kolme luonteeltaan hyvin erilaista konstruktioita tai konstruktiojoukkoa, joita kielitaustaltaan erilaiset suomenoppijat käyttivät eri tavoin. Analyysissa ei kuitenkaan erotu omaksi ryhmäkseen mitään yksittäistä kielen osa-aluetta tai yksittäistä ensikieltä, vaan erot vaikuttavat olevan luonteeltaan konstruktiokohtaisia. Yhtenä tämän tutkimuksen tavoitteena on arvioida keinoja, joilla ensikielikohtaisia eroja oppijankielessä ja sitä kautta kieltenvälisiä vaikutussuhteita voidaan tutkia aineisto- eli korpusvetoisesti. Tulokset eivät tarjoakaan selviä yksiselitteisiä todisteita ensikielen vaikutuksesta oppijansuomeen, vaan ennemminkin ne nostavat esiin sellaisia konstruktioita, joissa tällaisia vaikutuksia saattaa esiintyä. Sovellettu

menetelmä ja konstruktiolähestyminen mahdollistaakin potentiaalisten konstruktiokohtaisten, kieltenvälisistä historiallisista yhteyksistä tai laajemmasta samankaltaisuudesta riippumattomien vaikutuksen paikantamisen ilman etukäteen muodostettua hypoteesia. Tällä tavoin menetelmä mahdollistaa kohdennettujen kysymysten esittämisen käsillä olevista kielistä, mikä puolestaan helpottaa luonteeltaan epäselvien tai epäsuorien kieltenvälisen vaikutussuhteiden tutkimista ja tällaisten tutkimusasetelmien muodostamista.

Konjunktioita käytetään uncaria ja japania ensikielenä puhuvien suomenoppijoiden osa-aineistoissa likimain yhtä paljon kuin vertailuaineistossa ja selvästi enemmän kuin tšekkiä, liettuaa tai venäjää ensikielenä puhuvien suomenoppijoiden osa-aineistoissa. Yksikön ja sen ympäristön analyysi kuitenkin osoittaa, että konjunktioita käytetään kaikissa aineistoissa suhteellisesti samoissa määrin sekä esimerkin 28 kaltaisesti lauseita yhdistämässä että esimerkin 29 kaltaisesti lausekkeita yhdistämässä. Myös käytettyjen konjunktioiden leksikaalinen variaatio on eri osa-aineistoissa verraten samankaltaista, joskin yleisin konjunktio *ja* kattaa jonkin verran suuremman osan käytetyistä konjunktioista tšekkiä, liettuaa tai venäjää ensikielenä puhuvien osa-aineistoissa. Konjunktioiden osalta tulokset perustuvatkin lähinnä negatiiviseen evidenssiin: eron taustalla ei ole mikään yksittäinen lekseemi, laajempi konstruktio tai käyttöfunktio. Mahdollisia vaikuttavia tekijöitä saattavat olla esimerkiksi ensikielikohtaiset erot konjunktioiden käytössä (ks. Penttilä 2008) tai jotkin akateemista kirjoittamista koskevat konventionaaliset erot. Näiden seikkojen selvittäminen jää kuitenkin tulevan tutkimuksen varaan.

28) [...] miten ne idolit elävät *ja* mitä ne ostavat esim TV:ssa (S2-cs: las2-25tt01te05)

29) Vaikka välissä on *i ja e* [...] (S1: las2-vtt01vert128)

Yksikön kolmannen persoonan preesensmuotoista verbiä käytetään liettuaa ensikielenään puhuvien osa-aineistossa vähemmän kuin erityisesti uncaria ensikielenään puhuvien osa-aineistossa ja vertailuaineistossa. Yksikön ja ympäristön analyysi osoittaa, että suurin ero kyseisen muodon käytössä näiden osa-aineistojen välillä on perfektikonstruktion käytössä: liettuaa ensikielenään puhuvien aineistossa suhteellisesti suurempi osa preesensmuotoisten yksikön kolmannen persoonan verbeistä esiintyy esimerkin 30 kaltaisen preesensin sijaan esimerkin 31 tapaan osana perfektikonstruktiota.

30) ammatti *vaikuttaa* kovasti (S2-ja: las2-18tt01te10)

31) konjunktiiivista *on* kehittynyt *-isi-* (S2-lt: las2-24tt01te08)

Havainto ei luonnollisesti tarkoita, että perfekti olisi liettuaa ensikielenä puhuvien aineistossa yleisempi vaan ennen kaikkea sitä, että preesens saattaa olla aikamuotona tässä osa-aineistossa muita osa-aineistoja harvinaisempi, mistä syystä perfektikonstruktiot kattavat suuremman osan kaikista preesensmuotoisista verbeistä. Tämä saattaa puolestaan viitata ensikielen vaikutukseen aikamuotojen välisessä distribuutiossa ja tehtävänjaossa. Ilmiön tarkempi tutkimus vaatii kuitenkin tarkasteltavien kielten aikamuoto- ja aspektijärjestelmien kattavaa analyysia sekä kaikkien aikamuotojen tyypillisen käytön ja distribuution selvittämistä näissä kielissä.

Kolmas ensikielisiä eroja indikoiva konstruktio on merkitykseltään huomattavasti edellä kuvattuja spesifimpi ja käyttöalaltaan rajatumpi, esimerkkien 32 ja 33 kaltainen listauskonstruktio, joka voi puolestaan esiintyä nominaalisena osana muissa konstruktioissa, kuten esimerkissä 32 transitiivikonstruktiossa ja esimerkissä 33 intransitiivikonstruktiossa.

32) *Sanat järvi, lampi, meri ja joki* kuvaavat kaikki erilaisia vesistöjä (S1: las2-vtt01vert096)

33) *Pienet lähisukukielet vatja, vepsä, inkeroinen ja liivi* olivat kauan ilman kirjakieltä (S2-ru: las2-11tt01te08)

Tätä konstruktiota esiintyy aineistossa lähinnä venäjää ensikielenään puhuvien aineistossa sekä vertailuaineistossa. Konstruktiossa määritteet esiintyvät suomen yleisen sanajärjestyksen kannalta epäprototyypillisesti pääsanansa jälkeen, ja ne ovat nominatiivissa riippumatta pääsanansa sijamerkinnästä tai syntaktisesta funktiosta. Konstruktio noudattaa siis karkeasti skeemaa 'X:t/X:issä/X:ille/jne. X₁, X₂...X_n', ja sen keskeisenä funktiona on listata samanaikaisesti useita keskenään samankaltaisia määritteitä, mistä syystä esimerkiksi yhdyssana-konstruktion (ks. *järvi-, lampi-, meri- ja joki-sanat*) käyttäminen saattaa tuntua hankalalta. Koska määritteitä on monta, niiden pääsana sekä usein myös lauseen objekti tai predikatiivi ovat monikollisia. Rakenteen käyttöön saattaa vaikuttaa samanaikaisesti useita eri tekijöitä, kuten käsillä olevalle tekstilajille tyypillisten konventioiden tunteminen. Mahdollisen kieltenvälisen vaikutuksen vahvistaminen vaatisi tässäkin tapauksessa vastaavissa konteksteissa käytettävien konstruktioiden luonnetta informanttien ensikielissä, mutta on luontevaa ajatella tällaisten kohdekielen tyypillisestä rakenteesta poikkeavien konstruktioiden soveltamisen helpottuvan, mikäli kielenkäyttäjällä on aiempia kokemuksia rakenteeltaan samankaltaisesta konstruktioista.

5.4. Käyttöpohjainen näkökulma eroihin ja muutoksiin

Paneuduin tarkemmin kahteen sellaiseen kielen ilmiöön, joihin liittyvät konstruktiot nousivat esiin osana analyysia. Ensinnäkin modaalisuuden ilmaukset ja erityisesti mahdollisuutta ilmaisevat konstruktiot näyttävät erottavan suomea toisena kielenä kirjoittavien ja ensikielisten akateemista suomea kirjoittavien tekstejä toisistaan. Toiseksi presens- ja preteriti-aikamuotojen käyttö näyttää erottavan ensikielisiä ja suomea toisena kielenä kirjoittavia kielenkäyttäjiä toisistaan, minkä lisäksi aikamuotojen käyttö muuttuu huomattavasti tarkastelujakson aikana. Ilmiöiden tarkempi tarkastelu osoittaa, että molemmassa havainnoissa taustalla saattavat olla käsillä olevalle tekstilajille – kirjoitetulle akateemiselle suomelle ja sen alalajille tenttivastaukselle – erityiset konventiot, joita suomea toisena kielenä kirjoittavat kielenkäyttäjät eivät ole täysin oppineet – ainakaan tarkastelujakson alkuvaiheessa. Tämä puolestaan tukee ajatusta kielenkäytön, kielitaidon ja kielioppijärjestelmän perustavanlaatuisesta kontekstisidonnaisuudesta ja käyttöpohjaisesta perusolemuksesta. Ajatus vastaa pitkälti Bardovi-Harligin ja Dörnyein (1998: 234) huomautusta siitä, että oppijoiden kieliopin oppiminen ja kielellisten keinojen laadullinen kirjo eivät takaa samantasoista pragmaattisen tason kielitaitoa. Nyt tarkasteltavien kielen ilmiöiden erojen ja muutosten voidaan katsoa kuvastavan tällaista eroa: ilmiöön liittyvät konstruktiot kyllä hallitaan, mutta niiden tyypillinen käyttö ja erityisfunktiot käsillä olevassa, monille informanteille uudessa tai ainakin osin vieraassa tekstilajissa eivät välttämättä ole tuttuja. Näin syntyy kompleksisuusteorian termein uusi dynaaminen tilanne, joka näyttäytyy vertailuaineistoon nähden epätyypillisenä kielellisenä käyttäytymisenä ja johon yksilön kielijärjestelmä pyrkii reagoimaan mukautumalla käyttöpohjaisesti havaittuihin tyypillisyyksiin. Vastaavasti diskurssikaavojen (Östman 2005) näkökulmasta voidaan sanoa, että yksittäisten konstruktioiden tasolla muodon ja funktion väliset yhteydet hallitaan paremmin kuin diskurssin tasolla. Syynä saattaa olla se, että diskurssin tasolla halutun tekstin tyypilliset konstruktionaaliset piirteet eivät joko ole toisen kielen käyttäjille yhtä tuttuja tai ne voivat erota ensikielenään suomea käyttävistä kirjoittajista. Havaittu pitkittäismuutos tukee tulkintaa, sillä preteriti havaittiin sekä yhdeksi S1- ja S2-aineistoja erottavaksi avainkonstruktioksi että suurinta pitkittäismuutosta kuvastavaksi avainkonstruktioksi.

5.4.1. Mahdollisuuden ilmaukset

S1- ja S2-aineistoja erottavien avainmuotojen analyysissa selvisi, että A-infinitiivien (<inf1>) vähäisempi määrä S2-aineistossa juontuu modaalisen *voida*-verbin ja sitä seuraavan infinitiivin muodostaman verbiketjun käyttöeroista (ks. aiemmat esimerkit 21 ja 22 sivulla

69). Jotta voidaan selvittää, käyttävätkö suomea toisena kielenä kirjoittavat tämän sijaan ensikielisiä enemmän jotain muuta mahdollisuuden ilmausta vai ilmaistaanko suomea toisena kielenä kirjoittavien teksteissä todella vähemmän jonkin asian mahdollisuutta, on kuitenkin niin ikään tarkasteltava rinnakkain kaikkia suomen kielelle tyypillisiä mahdollisuuden ilmauksia. Mikäli mahdollisuutta todella ilmaistaan vähemmän, on lisäksi mielekästä selvittää, mikä tähän voisi olla syynä.

ISK:ssa (VISK §1551) mainittujen mahdollisuuden ilmausten rinnakkainen tarkastelu osoittaa, että mahdollisuutta ilmaistaan S2-aineistossa kaiken kaikkiaan likimain puolet S1-aineiston ilmauksien määrästä (S1: 18,9 / 1000 sanetta; S2: 10,6 / 1000 sanetta). Kummassakin osa-aineistossa noin kolme neljäsosaa mahdollisuuden ilmauksista on *voida*-verbistä ja jonkin verbin A-infinitiivistä koostuvia konstruktioita, eivätkä eri mahdollisuuden ilmausten jakaumat näin eroa osa-aineistojen välillä toisistaan. Mahdollisuuden ilmaukset siis ovat S2-aineistossa kaikkiaan S1-aineistoa harvinaisempia. Tarkastelusta jätettiin pois joitakin ISK:ssa (mp.) mainituista verbikonstruktioista niiden monimerkityksisyyden vuoksi. Tämä tuskin kuitenkaan vaikuttaa ilmiön kokonaiskuvaan, sillä kyseiset verbit eivät esiinny aineistossa A-infinitiivin yhteydessä juurikaan (ks. *S2-avainrakenteet*). Tulos on odotuksenmukainen myös aiemman tutkimuksen valossa. Kirjoittajan aseman ilmaisemisen suhteessa esitettyyn tietoon (engl. *stance*) on osoitettu olevan tyypillistä akateemiselle kirjoittamiselle (ks. esim. Campbell 1975; Hyland 1999). Modaalisten verbien on lisäksi osoitettu olevan ainakin akateemiselle englannille yksi tyypillisimpiä tapoja ilmaista tätä asemoitumista (Biber 2006: 95–103). Tällaisten tietyille tekstilajille tyypillisten piirteiden käytön on puolestaan havaittu erottavan ensikielisiä ja toisella kielellä toimivia kielenkäyttäjiä toisistaan (Mauranen 1996; Cotton 2010).

Mahdollisuuden ilmauksista yleisimmän – *voida*-verbin – tyypillinen käyttö tukee edellä kuvattua tulkintaa. Se on osa-aineistoissa likimain samankaltaista, mutta S1-aineistossa käytetään suhteellisesti enemmän passiivikonstruktioita. Lisäksi predikaatin semanttinen pääverbi kuvaa S1-aineistossa usein jonkinlaista luokittelua, kun tällaista tendenssiä ei ainakaan samassa määrin ole havaittavissa S2-aineistossa. Esimerkit 34 ja 35 kuvaavat passiivissa olevia, luokittelevia mahdollisuuden ilmauksia.

34) Suomessa verbit *voidaan jakaa* 0-, 1-, 2- tai 3-paikkaisiin verbeihin sen mukaan, moneenko tahoan ne vaikuttavat. [...] Suomen verbit *voidaan jakaa* myös hetkellistä ja pidempiaikaista toimintaa kuvaaviin verbeihin.
(S1: las2-vtt01vert010)

- 35) Mordvalainen eepos *voidaan luokitella* seuraaviin ryhmiin: 1) mytologinen ja 2) historiallinen. Mytologisissa eepoksissa aiheena on yleensä maailman luominen ja haltijoiden synty. [...] Historiallisissa eepoksissa kerrotaan mordvalaisten pitkästä taistelusta tataareja vastaan. [...] Eepokset ovat yleensä pitkiä, kuvat vaihtelevat nopeasti. (S2: las2-7tt01te11)

Aiempi tutkimus on osoittanut *voidaan*-sanamuodon olevan tyypillinen nimenomaan akateemiselle suomelle (Jantunen 2012), ja tiedon hierarkkisen jaottelun niin ikään ominaista tieteelliselle tyylille (Saukkonen 1984: 51–54). Lisäksi passiivin käytön on osoitettu alkavan vasta B-tasolla ja niiden käyttömäärän ja käytössä havaitun variaation lisääntyvän aina C-tasolle asti niin leksikon kuin käyttöfunktioidenkin kannalta (Seilonen 2013: 64–69). Tuntuukin hyvin uskottavalta, että mahdollisuuden ilmaisemisen ja sen tyypillisen käytön kaltaiset, käsillä olevan tekstilajin ominaispiirteet erottaisivat vielä edistyneitäkin suomenoppijoita ensikielisistä kielenkäyttäjistä. Käytettävät konstruktiot ovat oppijoille laadullisesti tuttuja, mutta niiden tyypillinen käyttö kirjoittajaposition ilmaisukeinona ei ole vielä vakiintunut ensikielisten kirjoittajien kaltaiseksi. Eri konstruktioiden välisten vaikutussuhteiden – kuten mahdollisuuden ilmausten ja passiivin – kontekstuaalisesti tyypillisen käytön kehkeytyessä käsillä olevan tekstilajin diskurssikaava on ainakin näiden kielenpiirteiden osalta vasta vakiintumassa. Tämä ajatus tuntuu perustellulta myös kompleksisuusteorian ja käyttöpohjaisen mallin kannalta: edistyneestä kielitaidosta huolimatta S2-aineiston informanteista valtaosa on vasta alkutaipaleella akateemisen suomen käytössä, minkä lisäksi kirjoitettu tenttivastaus saattaa tekstilajina olla ylipäättäänkin melko vieras. Näin kielijärjestelmässä on uusi dynaamisen muutoksen tila, jonka laukaisevana tekijänä on uusi tekstilaji (ks. Ellis & Larsen-Freeman 2006; Larsen-Freeman & Cameron 2008).

5.4.2. Aikamuodot muutoksessa

Kuten todettua, edistyneiden S2-oppijoiden aikamuotojen käyttö muuttuu kuudentoista kuukauden tarkastelujakson aikana radikaalisti niin, että preteritin käyttö puolittuu ja preesensia aletaan käyttää vastaavan määrän verran enemmän. Samanaikaisesti molempien aikamuotojen käyttöfrekvenssi lähenee S1-kirjoittajien aineistossa todettua tasoa. Kyseessä olevat muodot itsessään ovat kuitenkin varmasti laadullisesti tuttuja edistyneille oppijoille, minkä lisäksi mitkään tietyt sisäisen tai tyypillisen käyttöympäristön leksikaaliset, formaaliset tai funktionaaliset muutokset eivät selitä eroa. Sen sijaan on mahdollista, että havaittu frekvenssiero ja myöhempi muutos kuvastavat tekstilajikohtaista tyypillistä kielenkäyttöä,

jonka suomea toisena kielenä kirjoittavat informantit oppivat suomalaisessa yliopistossa tekemiensä opintojen edetessä.

Suomen preteriti on tyypillinen erityisesti luonteeltaan narratiivisille tekstilajeille, kuten kertomuksille ja raporteille, joissa tapahtumien sekventiaalinen järjestys on merkityksen kannalta keskeistä (VISK §1531). Akateemisessa kirjoitetussa suomessa keskiössä on kuitenkin useimmiten lukukokemuksen tai tutkimusprosessin kuvaamisen sijaan itse tieto ja tenttivastauksissa sitä kautta joko kurssilla opittu tai kirjasta luettu tietosisältö. Esimerkkien 36 ja 37 preteritin käytössä ei itessään ole mitään laadullisesti epäidiomaattista. Näin on laita valtaosassa aineiston preteritimuodoista. Itse lukukokemuksen kuvaaminen esimerkin 36 tapaan lähenee kuitenkin luonteeltaan narratiivista tyyliä, mitä voidaan pitää epätyypillisenä akateemiselle suomelle. Esimerkissä 37 keskiöön nousee preteritin vuoksi kirjoittajan kuvaaman tutkimuksen tutkimusprosessi sen sijaan, että teksti kohdistuisi tutkimuksen tuloksiin ja siitä saatuun tietoon. Tässä valossa preteritin käytön väheneminen saattaakin siis olla tekstien narratiivisen luonteen vähenemistä ja S2-oppijoiden tekstilajille ominaisten toimintatapojen oppimista (ks. Hyland 2004: 54).

- 36) Olen saanut artikkeleista vinkkejä siitä, miten on mahdollista opettaa eri lausetyyppejä, koska artikkeleissa *oli* esillä kokonainen järjestelmä siitä, miten eri lausetyypit on esitettävä. (las2-21tt01te06)
- 37) Suurin osa tehtävistä *oli* lapsille vaikea, mutta vaikeimmat *olivat* sanojen tavun poistuminen ja lisääminen. Kun lapset *lisäsivät* tavut sanoihin, jotkin sanat olivat vaihtunut toisiksi sanoiksi. Sanojen tavun poistumisessa tehtävässä *oli* myös nähtävissä samoin. (las2-12tt01te08)

Muutoksen – eli tässä tapauksessa kontekstuaalisesti konventionaalisen kielenkäytön oppimisen – voi siis nähdä käynnistyvän käyttöpohjaisesti, kun oppijat orientoituvat ja tutustuvat itselleen osin vieraaseen tekstilajiin eli akateemiseen kirjoitettuun suomeen ja sen alalajiin tenttivastaukseen. On mahdollista, että tarkastelujakson alun preteritien runsas määrä liittyy tekstien dynaamiseen tyyliin (ks. Andersen & Shirai 1994), ja tuntuukin todennäköiseltä, että preteritin vähenevä käyttö liittyy ennemmin tekstilajikohtaisten diskurssikaavojen (ks. Östman 2005) oppimiseen kuin muutoksiin kielenkäyttäjien aikamuoto- ja aspektijärjestelmässä. Uusi tekstilaji laukaisee funktioiden uudelleenjärjestymisen kohti käsillä olevalle tekstilajille tyypillistä käyttöä – oppimisprosessi kuvastaa siis luonteeltaan kielijärjestelmän ja kielenoppimisen rakenteellista analogiaa. Kielijärjestelmän osa – tässä tapauksessa tempusjärjestelmän funktionaalinen distribuutio –

siirtyy kielellisessä syötöksessä ja kielenulkoisessa kokemusympäristössä tapahtuvan muutoksen vuoksi dynaamiseen tilaan, ja järjestelmä muuttuu tältä osin merkityksellisten diskurssikaavojen paremman hallinnan myötä kohti ensikielistä kielenkäyttöä.

6. YHTEENVETO JA PÄÄTELMÄT

Olen tässä väitöskirjatutkimuksessa tutkinut sitä, miten oppijankielen rakenteellisia ominaispiirteitä voidaan tutkia aineistolähtöisesti ilman ennalta tehtyjä oletuksia mahdollisten erojen luonteesta, sitä, minkälaiset konstruktiot erottavat edistyneiden suomenoppijoiden tekstejä ensikielisten suomenpuhujien teksteistä, sekä sitä, mistä havaitut rakenteelliset erot voivat johtua. Yhteenveto tuloksista on luettavissa luvussa 5. Tulosten perusteella vastaan esittämiini tutkimuskysymyksiin seuraavasti:

1) Miten oppijankielen rakenteellisia ominaispiirteitä voidaan tunnistaa ja analysoida korpus- eli aineistovetoisesti – ilman ennalta asetettua hypoteesia mahdollisten erojen luonteesta? Tässä tutkimuksessa kehitetty ja käytetty menetelmä avainrakenneanalyysi soveltuu korpusvetoiseen oppijankielen rakenteen tutkimukseen. Menetelmä soveltaa avainsana-analyysistä tuttua tilastollisen avaimuuden käsitettä kontrastiiviseen välikieliansalyysiin, ja se hyödyntää morfosyntaktisesti annotoidun kielikorpuksen tarjoamia mahdollisuuksia yhdistäen yksikön ja sen ympäristön analyysin konstruktiokielioppiin ja sitä kautta laajemmin käyttö pohjaiseen kielioppikäsitukseen. Lähtökohtana olivat sanojen morfologisten muotojen käytössä havaitut frekvenssierot, mutta menetelmän avulla havaittujen määrällisten erojen kautta päästään käsiksi luonteeltaan monenlaisiin muodon ja merkityksen toistuviin yhdistelmiin – konstruktioihin – joiden käyttö erottaa vielä edistyneitäkin suomenoppijoita ensikielistä kielenkäyttäjistä. Havaitut yhdistelmät vaihtelevat *esimerkiksi*-sanamuodon kaltaisista, leksikaalisesti spesifeistä konstruktioista modaalisten verbiketjujen kaltaisiin abstrakteihin konstruktioihin. Menetelmä mahdollistaa myös ensikielikohtaisten erojen ja käynnissä olevan välikielen muutosprosessin tavoittamisen ja tunnistamisen aineistovetoisesti.

2) Mitkä kielen konstruktiot ovat tyypillisiä tai epätyypillisiä edistyneiden suomenoppijoiden kirjoituksille ensikielenään suomea käyttävien kirjoituksiin verrattaessa? Edistyneitä suomenoppijoita ensikielistä kirjoittajista erottavia konstruktioita on monenlaisia. Erot liittyvät usein oppijansuomen vähäisempään morfosyntaktiseen kompleksisuuteen (esim. VA-partisiipista muodostettu referatiivirakenne) tai vähäisempään modaalisuuden ilmaistamiseen (esim. modaaliset verbiketjut), joidenkin käsillä olevalle tekstilajille tyypillisten leksikaalisesti rajattujen konstruktioiden vähäisempään käyttöön oppijansuomessa (esim. sana *esimerkiksi*) tai aikamuotojen erilaiseen käyttöön (preteritin runsaampi käyttö oppijansuomessa). Jotkut havaituista eroista ovat

sellaisia, että niitä on käsitelty S2-tutkimuksessa jo aiemmin, kun taas toiset tuovat uutta tietoa tyypillisistä eroista. Tässä tutkimuksessa tehdyt päätelmät eivät perustu ennalta muodostettuihin hypoteeseihin mahdollisista eroista, vaan lähtokohtana ovat aineistovetoiset havainnot aineistojen välisistä eroista. Näin ollen havainnot valottavat myös jo tutkittuja piirteitä osin uudesta näkökulmasta. Aiemmin tutkittujen piirteiden nouseminen esiin kuitenkin tukee menetelmän käyttökelpoisuutta tyypillisten erojen paikantajana ja parantaa tulosten luotettavuutta myös niiden konstruktioiden osalta, joita ei ole aiemmin tarkasteltu S2-näkökulmasta.

3) Mitkä kielen konstruktiot erottavat eri ensikieliä puhuvia edistyneitä suomenoppijoita toisistaan? Unkaria, japania ensikielenään puhuvat informantit käyttävät konjunktioita muita suomenoppijaryhmiä enemmän ja samankaltaisesti ensikielenään suomea puhuvien kanssa. Liettuaa ensikielenään puhuvat informantit puolestaan käyttävät preesensmuotoisia verbejä unkaria ensikielenään puhuvia informantteja ja ensikieleltään suomalaisia kielenkäyttäjiä vähemmän. Listauskonstruktiota 'X:t/X:issä X₁, X₂...X_n' puolestaan käytetään lähes yksinomaan venäjää ensikielenään puhuvien aineistossa sekä vertailuaineistossa. Tutkimuksessa ei tarkasteltu, minkälaiset kieltenväliset vaikutussuhteet saattavat synnyttää havaittuja eroja, mutta tulosten avulla jatkotutkimus voidaan kohdistaa tarkemmin tiettyihin konstruktioihin.

4) Minkä kielen konstruktioiden käyttö muuttuu edistyneiden suomenoppijoiden kirjoituksissa ajan kuluessa ja miltä muutos näyttää ensikielenään suomea käyttävien kirjoituksiin verrattuna? Edistyneiden suomenoppijoiden preteritin käyttö vähenee kuudentoista kuukauden tarkastelujakson aikana ja samanaikaisesti preesensin käyttö lisääntyy vastaavissa määrin. Määrällisesti käyttö lähenee ensikielenään suomea puhuvien vertailuaineistoa, mutta laadullisesti käytössä on eroja vielä tarkastelujakson lopussakin.

5) Mistä syistä vielä edistyneidenkin suomenoppijoiden kieli poikkeaa rakenteellisesti joiltakin osin ensikielenään suomea käyttävien kielestä ja on edelleen muutoksessa? Monet havaituista eroista osoittavat, että ensikielisiä ja toisella kielellä kirjoitettavia edistyneitä kielenoppijoita erottavat nimenomaan käsillä olevan tekstilajin kielelliset erityispiirteet ja tyypilliset konventiot ja näissä havaitut epätyypilliset käyttömäärät ja -tavat – ei niinkään se, että oppijat eivät laadullisesti osaisi muodostaa tai käyttää havaittuja konstruktioita. Kun kielen ja sen konstruktioiden oppimista ja käyttöä kuvataan distribuutiolähtöisesti – eli kun

kuvataan sitä, kuinka paljon ja millä tavoin joitakin kielen rakenteita käytetään – ei siis voida sivuuttaa käyttökontekstin merkitystä eli erilaisten diskurssikaavojen hallintaa. Esimerkiksi erot preteritin käytössä ja asiointilan mahdollisuuden runsas ilmaiseminen ensikielisten teksteissä näyttävät liittyvän akateemiselle kirjoittamiselle tyypilliseen kielenkäyttöön – mahdollisuuden ilmauksia käytetään kirjoittajan mielipiteen etäännyttämiseen sanotusta, kun taas suomea toisena kielenä kirjoittavien suosima preteriti korostaa akateemiselle kirjoittamiselle epätyypillistä narratiivisuutta, joka korostaa kirjoittajan luku- ja oppimiskokemusta käsitellyn tiedon kustannuksella.

Tämän tutkimuksen tulokset puoltavat kielitaidon ja kielenkäytön kontekstisidonnaisuutta ja sitä kautta kielen oppimisen pohjimmiltaan käyttöpohjaista luonnetta – kontekstuaalisesti idiomaattinen kielenkäyttö ja täysipainoinen kyky toimia kyseisessä tekstilajissa vaatii altistumista kyseiselle käyttökontekstille ja sen tyypillisille kielellisille käytänteille, vaikka kielenkäyttäjän kielitaito ja kielellinen valikoima laadullisesti kattaisikin kaikki tarvittavat keinot. Kielenoppijan kielijärjestelmä muuttuu tai ainakin se voi muuttua todellisen kielenkäytön ja käyttötarpeiden mukaan ja niiden seurauksena. Konstruktioiden oppiminen ja niiden käyttö siis perustuvat idiomiperiaatteen mukaisesti todelliseen kielenkäyttöön myös oppijankielessä, mutta tekstilajin osittaisen vierauden eli kielellisen syötöksen erilaisuuden vuoksi oppijat käyttävät kieltä osin eri tavalla kuin ensikieliset kielenkäyttäjät.

Tuntuu luontevalta ajatella, että avainrakenneanalyysia menetelmänä ja vastaavankaltaista tutkimusasetelmaa voitaisiin soveltaa oppijankielen tutkimuksen ohella myös muunlaisiin tutkimuksiin. Oppijankielen tutkimuksen kannalta olisi mielenkiintoista vertailla vastaavalla tavalla rinnakkain samanaikaisesti useita keskenään erilaisia tekstilajeja, niissä havaittavia kielenoppijoiden ja ensikielisten kielenkäyttäjien välisiä eroja, eri tekstilajien välisiä eroja sekä näiden välisiä suhteita toisiinsa. Menetelmää voidaan soveltaa (ja on jossain määrin jo sovellettukin, esim. Ivaska & Siitonen 2015) eri kielitaitotason aineistoihin sillä edellytyksellä, että aineistojen tekstilajikohtainen luonne on selvillä ja voidaan ottaa huomioon eroja ja yhtäläisyyksiä analysoitaessa. Menetelmän soveltaminen tekstilajien tutkimukseen ylipäätään tuntuu niin ikään mielekkäältä, samoin mahdollisten diakronisten muutosten tarkasteleminen.

Vaikka tässä väitöskirjatutkimuksessa saadut tulokset onkin tavoitettu korpusvetoisesti, vaatii määrällisten havaintojen tulkinta myös huolellista laadullista analyysia. Menetelmä ei

tarjoa tutkijalle valmiita avainrakenteita eli muoto-merkitys-pareja, jotka olisivat tyypillisiä tai epätyypillisiä oppijansuomelle vaan ainoastaan frekvenssieron perusteella havaittuja avainmuotoja, jotka saattavat indikoida eroa. Näin ollen konstruktionaalinen tulkinta on tarkemman analyysin ja tätä kautta väistämättä myös tutkijan tulkinnallisten ratkaisujen varassa. Menetelmä ei myöskään mahdollista suoraan esimerkiksi tekstin viittausrakenteen kaltaisten, kontekstuaalisten taustamuuttujien tai yksilöiden tarkempien kielenulkoisten, oppijakohtaisten muuttujien huomioonottamista tulkittaessa konstruktioiden käyttöerojen syitä. Lisäksi käsillä olevan tekstilajin tuntemusta tai sille altistumista ei kykene eksplisiittisesti havainnoimaan nyt käytetyn aineiston ja menetelmän keinoin. Käsillä oleva aineisto rajoittaa lisäksi menetelmän sovellettavuutta, sillä analyysin kohdistaminen yksittäisiin morfologisiin piirteisiin ei ole mahdollista käsillä olevan kaltaisessa aineistossa, jossa annotaatio on kohdistettu vain saneiden tarkkuudella. Morfeemeittain kohdistettu korpus mahdollistaisi yksittäisten avainmorfeemien paikantamisen, mikä tuntuu suomen kaltaisten, morfologisesti rikkaiden kielten kannalta mielekkäältä lähestymistavalta.

Tenttivastausten kaltaisille teksteille keskeisen tehtävänannon vaikutusta ei ole tässä tutkimuksessa tarkasteltu eksplisiittisesti, sillä isoa osaa tehtävänannoista ei ole tallennettu korpukseen. Tehtävänantojen ja tenttivastausten suhteen tarkastelu saattaisi hyvinkin selittää joitain tässäkin tutkimuksessa tarkastelluista seikoista, ovathan näiden kahden tekstilajin väliset suhteet poikkeuksellisen vahvat (ks. esim. Makkonen-Graig 2008: 213). Toisaalta tehtävänantojen käyttö saattaa riippua kielenkäyttäjän kielitaitotasosta (ks. Plakans & Gebril 2013). Korpusvetoisuudesta huolimatta monia tulkintoja ohjaavat aineiston lisäksi väistämättä myös tulkitsijan omat aiemmat kokemukset ja havainnot, ja eri tutkija saattaisi päätyä joiltakin osin erilaisiin tulkintoihin. Menetelmän jatkokehittämissä onkin syytä pohtia myös niitä tapoja, joilla tehdyt havainnot voidaan yhtäältä linkittää abstraktimpiin tekstuaalisiin piirteisiin ja toisaalta oppijakohtaisiin tekijöihin.

Monet nyt havaitut erot ansaitsevat syvempää paneutumista niihin tekijöihin, jotka vaikuttavat ilmiön taustalla, ja esimerkiksi kieltenvälisen vaikutussuhteiden tutkimus jää tässä tutkimuksessa lähinnä potentiaalisten konstruktionaalisten erojen toteamiseen. Tämän tutkimuksen menetelmä ja sillä saadut tulokset eivät siis itsessään osoita kieltenvälisiä vaikutussuhteita, eivätkä ne näin ollen kuvaa suoraan sitä, miten jokin aiemmin opittu kieli vaikuttaa oppijansuomeen. Tutkimuksessa käytetyt kriteerit olivat verraten tiukat, ja on mahdollista, että tarkastelun ulkopuolelle rajautui kieltenvälisen vaikutussuhteiden kannalta

mielenkiintoisia konstruktioita. Toisaalta tiukat kriteerit takaavat sen, että tehdyt havainnot hyvin todennäköisesti liittyvät nimenomaan kieltenvälisiin vaikutussuhteisiin. Kaikkiaan menetelmä helpottaa kieltenvälisen vaikutusten kannalta potentiaalisesti relevanttien piirteiden paikantamista ja tätä kautta tarkempien tutkimuskysymysten ja -asetelmien muotoilemista. Menetelmän jatkokehittämissä onkin tärkeää pohtia, miten aiemmin opittujen kielten konstruktioita voitaisiin linkittää havaittuihin ensikielikohtaisiin eroihin.

Menetelmän kehittämissä kyseessä on nähdäkseni tietynlainen vaihtokauppa-tilanne: mitä enemmän muuttujia otetaan huomioon jonkin kielenoppimisen kannalta mielenkiintoisen ilmiön tulkitsemisessa, sitä tarkempia ja varmempia tulkintoja ilmiön syistä voidaan esittää. Samanaikaisesti kutistuu kuitenkin tutkimuksen tulkinnallinen avaruus, ja näin laajemman kielellisen kokonaistilanteen huomioiminen vaikeutuu. Toisin sanoen kyse on ainakin osin siitä, halutaanko tutkimuksessa korostaa sen tulkinnallista syvyyttä vai leveyttä. Korpusvetoisuuden kiistämättömänä etuna on se, että tutkimuksen keskiöön tulee nostettua ilmiöitä, jotka muuten saattaisivat jäädä tarkastelun ulkopuolelle. Lähestymistapojen yhdistäminen on nähdäkseni hedelmällinen maaperä uudelle tutkimukselle: avainrakennepiirteiden kaltaiset korpusvetoiset lähestymistavat voivat tarjota tutkijayhteisölle tutkimuksellisia syöttejä, joihin muunlaiset lähestymistavat voivat tarttua ja paneutua perusteellisemmin eri näkökulmista ja erilaisia menetelmiä hyväksi käyttäen. Kompleksisuusteoriaa mukaillen: kielen oppiminen on äärimmäisen monimutkainen ja monimuuttujainen ilmiö, ja se ansaitsee tulla tutkituksi ja tulkituksi monista eri näkökulmista monia erilaisia kysymyksenasetteluja ja menetelmiä hyödyntäen.

LÄHTEET

- Aarts, Jaan & Granger, Sylviane 1998: Tag sequences in learner corpora: a key to interlanguage grammar and discourse. – Sylviane Granger (toim.), *Learner English on Computer* s. 132–141. London: Longman.
- Atkins, Sue, Clear, Jeremy & Ostler, Nicholas 1992: Corpus Design Criteria. – *Literary and Linguistic Computing* 7(1) s. 1–16.
- Bakhtin [Bahtin], Mihail M. 1986 [1953]: The problem of speech genres. Kääntänyt Vern W. McGee. – Caryl Emerson & Michael Holquist (toim.), *Speech genres and other late essays* s. 60–102. Austin: University of Texas Press.
- Bardovi-Harlig, Kathleen 2000: *Tense and Aspect in Second Language Acquisition: Form, Meaning, and Use*. Oxford: Benjamins.
- Bardovi-Harlig, Kathleen & Dörnyei, Zoltán 1998: Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. – *TESOL Quarterly* 32(2) s. 233–259.
- Bardovi-Harlig, Kathleen & Hartford, Beverly S. 1990: Congruence in native and nonnative conversations: Status balance in the academic advising session. – *Language Learning* 40(4) s. 467–501.
- Bardovi-Harlig, Kathleen & Hartford, Beverly S. 1993: Learning the rules of academic talk: A longitudinal study of pragmatic development. – *Studies in Second Language Acquisition* 15(3) s. 279–304.
- Barlow, Michael & Kemmer, Suzanne (toim.) 2000: *Usage based Models of Language*. Chicago: CSLI Publications.
- Bates, Elizabeth & Goodman, Judith C. 1997: On the inseparability of grammar and lexicon: Evidence from acquisition, aphasia and real-time processing. – *Language and Cognitive Processes* 12(5/6) s. 507–584.
- Bates, Elizabeth & MacWhinney, Brian 1987: Competition, variation and language learning. – Brian MacWhinney (toim.), *Mechanisms of Language Acquisition* s. 157–193. Hillsdale (NJ): Lawrence Erlbaum Associates.
- Bhatia, Vijay K. 1993: *Analysing genre: Language use in professional settings*. London: Longman.
- Belcher, Diane & Braine, George (toim.) 1995. *Academic Writing in a Second Language: Essays on Research and Pedagogy*. Norwood (NJ): Ablex Publishing Corporation.
- Bencini, Giulia & Goldberg, Adele, E. 2000: The contribution of argument structure constructions to sentence meaning. – *Journal of Memory and Language* 43(4) s. 640–651.
- Berkenkotter, Carol & Huckin, Thomas N. 1995: *Genre knowledge in disciplinary communication*. Hillsdale (NJ): Lawrence Erlbaum.
- Biau, Gérard 2012: Analysis of a random forests model. – *Journal of Machine Learning Research* 13 s. 1063–1095.
- Biber, Douglas 1988: *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas 1993: Representativeness in corpus design. – *Literary and Linguistics Computing* 8(4) s. 243–257.
- Biber, Douglas 2006. *University Language. A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.
- Biber, Douglas & Conrad, Susan 1999: Lexical bundles in conversation and academic prose. – Hilde Hasselgård & Signe Oksefjell (toim.), *Out of Corpora: Studies in Honor of Stig Johansson* s. 181–189. Amsterdam: Rodopi.
- Bley-Vroman, Rorber 1983: The comparative fallacy in interlanguage studies: The case of systematicity. – *Language Learning* 33(1) s. 1–17.

- Blommaert, Jan & Rampton, Ben 2011: Language and superdiversity. – *Diversities* 13(2) s. 1–21.
- Bohn, Ocke-Schwen & Flege, James 1992: The production of new and similar vowels by adults German learners of English. – *Studies in Second Language Acquisition* 14(2) s. 131–158.
- Brattico, Pauli 2013: Introspektio, kielen autonomia ja aineistot: pohdintoja kielitieteen menetelmistä. *Virittäjä* 117 s. 102–109.
- Breiman, Leo 2001: Random forests. – *Machine Learning* 45 (1) s. 5–32.
- Bybee, Joan 1985: *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.
- Bybee, Joan 2010: *Language, Usage, and Cognition*. Cambridge University Press, Cambridge.
- Bybee, Joan & Hopper, Paul 2001: Introduction to frequency and emergence of linguistic structure. – Joan Bybee & Paul Hopper (toim.), *Frequency and the Emergence of Linguistic Structure* s. 1–24. Amsterdam: John Benjamins.
- Bybee, Joan & Thompson, Sandra 1997: Three frequency effects in syntax. – *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Pragmatics and Grammatical Structure* s. 65–85.
- Campbell, Paul N. 1975. The personae of scientific discourse. – *Quarterly Journal of Speech* 61(4) s. 391–405.
- CEFLING = Linguistic Basis of the Common European Framework for L2 English and L2 Finnish. Tutkimushanke Jyväskylän yliopistossa 2007–2009, hankkeen verkkosivut. <https://www.jyu.fi/hum/laitokset/kieliet/cefling/suom> (30.8.2012).
- CEF = Eurooppalainen viitekehys: kielten oppiminen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys 2003. WSOY, Helsinki.
- Celce-Murcia, Marianne, Dörnyei, Zoltán & Thurrell, Sarah 1995: Communicative competence: A pedagogically motivated model with content specifications. – *Issues in Applied Linguistics* 6(2) s. 5–35.
- Cenoz, Jasone, Hufeisen, Britta & Jessner, Ulrika (toim.) 2001: *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Bilingual Education and Bilingualism 31. Clevedon: Multilingual Matters.
- Chen, Winnie, Greaves, Chris & Warren, Martin 2006: From n-gram to skipgram to concgram. – *International Journal of Corpus Linguistics* 11(4) s. 411–433.
- Chomsky, Noam 1965: *Aspects of the Theory of Syntax*. Massachusetts: MIT Press.
- Cotton, Fiona 2010. Critical thinking and evaluative language use in academic writing: A comparative cross-cultural study. – George Blue (toim.), *Developing Academic Literacy* s. 73–85. Oxford: Peter Lang.
- Cowie, Anthony, Paul 1981: The treatment of collocations and idioms in learners' dictionaries. – *Applied Linguistics* 2(3) s. 223–235.
- Croft, William 2001: *Radical Construction Grammar - Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Croft, William & Cruse, D. Alan 2004: *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Dagneaux, Estelle, Denness, Sharon & Granger, Sylviane 1998: Computer-aided error analysis. – *System: An International Journal of Educational Technology and Applied Linguistics* 26(2) s. 163–174.
- Devitt, Amy J. 2004: *Writing Genres*. Carbondale: Southern Illinois University Press.
- Dörnyei, Zoltán 2005: *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah (NJ): Lawrence Erlbaum.
- Dörnyei, Zoltán, & Skehan, Peter 2003. Individual differences in second language learning. – Catherine J. Doughty & Michael. H. Long (toim.), *The handbook of second language acquisition* s. 589–630. Oxford: Blackwell.

- Eddington, David 2008: Linguistics and the scientific method. – *Southwest Journal of Linguistics* 27 s. 1–16.
- Ellis, Nick C. 1998: Emergentism, connectionism, and language learning. – *Language Learning* 48(4) s. 631–664.
- Ellis, Nick C. 2002: Frequency effects in language processing and acquisition. A review with implications for theories of implicit and explicit language acquisition. – *Studies in Second Language Acquisition* 24(2) s. 143–188.
- Ellis, Nick C. 2003: Constructions, chunking, and connectionism: The emergence of second language structure. – Catherine J. Doughty & Michael H. Long (toim.), *The Handbook of Second Language Acquisition* s. 63–103. Malden (MA): Blackwell Publishing.
- Ellis, Nick C. 2005: At the interface: dynamic interactions of explicit and implicit knowledge. – *Studies in Second Language Acquisition* 27(2) s. 305–352.
- Ellis, Nick C. 2006: Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking and perceptual learning. – *Applied Linguistics* 27(2) s. 164–194.
- Ellis, Nick C. 2008: Phraseology: The periphery and the heart of language. – Fanny Meunier & Sylviane Granger (toim.), *Phraseology in Foreign Language Learning and Teaching* s. 1–14. Amsterdam: John Benjamins.
- Ellis, Nick C. & Larsen-Freeman, Diane 2006: Language emergence: Implications for applied linguistics. – *Applied Linguistics* 27(4) s. 558–589.
- Ellis, Rod 2004: Individual differences in language learning. – Catherine Elder & Alan Davies (toim.), *Handbook of Applied Linguistics* s. 525–551. Oxford: Blackwell.
- Ellis, Rod 2008: *The Study of Second Language Acquisition*. Toinen, korjattu ja lisätty painos. Oxford: Oxford University Press.
- Elman, Jeffrey L. 1990: Finding structure in time. – *Cognitive Science* 14(2) s. 179–211.
- Fillmore, Charles J. 1975: An alternative to checklist theories of meaning. – *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society* s. 123–131.
- Fillmore, Charles J. 1977: Topics in lexical semantics. – Roger W. Cole (toim.), *Current Issues in Linguistic Theory* s. 76–138. Bloomington: Indiana University Press.
- Fillmore, Charles J. 1985: Syntactic intrusions and the notion of grammatical construction. *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*, s. 73–86.
- Fillmore, Charles J., Kay, Paul & O'Connor, Mary Catherine 1988: Regularity and idiomatity in grammatical constructions: The case on let alone. – *Language* 64(3) s. 501–538.
- Firth, John Rupert 1968: *Selected Papers of J. R. Firth 1952–1959*. – Frank Robert Palmer (toim.). London: Longman.
- Francis, Gill 1993: A corpus-driven approach to grammar – principles, methods and examples. – Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and Technology. In Honour of John Sinclair* s. 137–156. Amsterdam: John Benjamins.
- Garside, Roger & Leech, Geoffrey & McEnery, Tony (toim.) 1997. *Corpus Annotation: Linguistic Information from Computer Text Corpora*. New York: Longman.
- Goldberg, Adele E. 1995: *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Goldberg, Adele E. 2006: *Constructions at Work – The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele E. & Casenheiser, Devin 2008: Construction learning and second language acquisition. – Peter Robinson & Nick C. Ellis (toim.), *Cognitive Linguistics and Second Language Acquisition* s. 197–215. Amsterdam: John Benjamins.
- Goldberg, Adele E. & Casenheiser, Devin & Sethuraman, Nitya 2004: Learning argument structure generalizations. – *Cognitive Linguistics* 15(3), s. 289–316.

- Granger, Sylviane 1996: From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. – Karin Aijmer, Bengt Altenberg & Mats Johansson (toim.), *Languages in Contrast* s. 37–51. Lund: Lund University Press.
- Granger, Sylviane 1998: Prefabricated patterns in advanced EFL writing: collocations and formulae. – Anthony Paul Cowie (toim.), *Phraseology: theory, analysis and applications* s. 145–160. Oxford: Oxford University Press.
- Granger, Sylviane 2002: A bird's-eye view of learner corpus research. – Sylviane Granger, Joseph Hung & Stephanie Petch-Tyson (toim.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* s. 3–33. Language Learning and Language Teaching 6. Philadelphia: John Benjamins.
- Granger, Sylviane 2008: Learner corpora. – Anke Lüdeling & Merja Kytö (toim.), *Corpus Linguistics. An International Handbook*, vol 1 s. 259–275. Handbooks of Linguistics and Communication Science 29.1. Berlin: Mouton de Gruyter.
- Granger, Sylviane 2015: Contrastive Interlanguage Analysis: A Reappraisal. – *International Journal of Learner Corpus Research* 1(1) s. 7–24.
- Granger, Sylviane & Paquot, Magali 2008: Disentangling the phraseological web. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An Interdisciplinary Perspective* s. 27–49. Amsterdam: John Benjamins.
- Gries, Stefan Th. 2008: Phraseology and linguistic theory. – Sylviane Granger & Fanny Meunier (toim.), *Phraseology: An Interdisciplinary Perspective* s. 3–25. Amsterdam: John Benjamins.
- Gries, Stefan Th. & Adelman, Allison S. 2014: Subject realization in Japanese conversation by native and non-native speakers: exemplifying a new paradigm in learner corpus research. – Jesús Romero-Trillo (toim.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*. Yearbook of Corpus Linguistics and Pragmatics 2. New York: Springer.
- Gries, Stefan Th. & Deshors, Sandra C. 2014: Using regressions to explore deviations between corpus data and a standard/target: two suggestions. – *Corpora* 9(1) s. 109–136.
- Gries, Stefan Th. & Hilpert Martin 2008: The identification of stages in diachronic data: variability-based neighbor clustering. – *Corpora* 3(1) s. 59–81.
- Gries, Stefan Th. & Stefanowitsch, Anatol 2004: Extending collocation analysis: A corpus-based perspectives on 'alternations'. – *International Journal of Corpus Linguistics* 9(1) s. 97–129.
- Gries, Stefan Th. & Wulff, Stefanie 2005: Do foreign language learners also have constructions? Evidence from priming, sorting and corpora. – Francisco José Ruis de Mendoza Ibáñez (toim.), *Annual Review of Cognitive Linguistics* 3 s. 182–200. Amsterdam: John Benjamins.
- Gries, Stefan Th. & Wulff, Stefanie 2009: Psycholinguistic and corpus-linguistic evidence for L2 constructions. – Francisco José Ruis de Mendoza Ibáñez (toim.), *Annual Review of Cognitive Linguistics* 7 s. 163–186. Amsterdam: John Benjamins.
- Guthrie, David, Allison, Ben, Liu, Wei, Guthrie, Louise & Wilks, Yorick 2006: A closer look at skip-gram modelling. – *Proceedings of Fifth international Conference on Language Resources and Evaluation (LREC), Genoa, Italy* s. 1222–1225. <http://www.lrec-conf.org/proceedings/lrec2006/> (06.03.2012).
- Hakulinen, Auli & Leino, Pentti 2006: Genre fennistiikassa. – Mickael Suominen, Antti Arppe, Anu Airola, Orvokki Heinämäki, Matti Miestamo, Urho Määttä, Jussi Niemi, Kari K. Pitkänen & Kaius Sinnemäki (toim.), *A Man of Measure. Festschrift in Honour of Fred Karlsson on his 60th Birthday. Special Supplement to SKY Journal of Linguistics* vol. 19 s. 12–23. Turku: The Linguistic Association of Finland.
- Halliday, Michael, A. K. 2004: *An Introduction to Functional Grammar*. Kolmas, korjattu ja muokattu painos. Toim. Christian M. I. M. Matthiessen. London: Hodder Education.

- Halliday, Michael, A. K. & Hasan Ruqaiya 1976: *Cohesion in English*. London: Longman.
- Halliday, Michael, A. K. & Hasan, Ruqaiya 1989: *Language, context and text: Aspects of language in a social semiotic perspective*. Oxford: Oxford University Press.
- Hamunen, Markus 2015: Analogian käsitteestä konstruktiokieliopissa. – *Virittäjä* 119 s. 278–286.
- Hardie, Andrew & McEnery, Tony 2010: On two traditions in corpus linguistics, and what they have in common. – *International Journal of Corpus Linguistics* 15(3) s. 384–394.
- Hashemi, Mohammad R. 2012: Reflections on mixing methods in applied linguistics research. – *Applied Linguistics* 33(2) s. 206–212.
- Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.) 2012: *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus.
- Hilpert, Martin 2006. Distinctive collexeme analysis and diachrony. – *Corpus Linguistics and Linguistic Theory* 2(2) s. 243–257.
- Hoey, Michael 1997: From concordance to text structure: new uses for computer corpora. – Barbara Lewandowska-Tomaszczyk & P. Melia (toim.), *PALC '97: Proceedings of Practical Applications of Linguistic Corpora Conference, University of Lodz* s. 2–23.
- Hollander, Myles & Wolfe, Douglas A. 1973: *Nonparametric statistical methods*. New York: John Wiley & Sons.
- Hopper, Paul J. 1987: *Emergent grammar*. – *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* s. 139–157.
- Hulstijn, Jan 2002: Towards a unified account of the representation, processing and acquisition of second language knowledge. – *Second Language Research* 18(3) s. 193–223.
- Hunston, Susan 2008: Collection strategies and design decisions. – Anke Lüdeling & Merja Kytö (toim.), *Corpus Linguistics. An International Handbook*, vol 1 s. 154–168. Handbooks of Linguistics and Communication Science 29.1. Berlin: Mouton de Gruyter.
- Hunston, Susan & Francis, Gill 2000: *Pattern Grammar. A Corpus-Driven approach to the Lexical Grammar of English*. Studies in Corpus Linguistics 4. Amsterdam: John Benjamins.
- Huumo, Tuomas 2007: Joko lingvistin nojatuoli joutaisi kaatopaikalle? Introspektiolingvistikän asemasta kognitiivisessa kielitieteessä. – *Emakeele seltsiraamat* 53 s. 163–180.
- Hyland, Ken 1999: Disciplinary discourses: Writer stance in research articles. – Christopher N. Candlin & Ken Hyland (toim.), *Writing: Texts, Processes and Practices*. London: Longman.
- Hyland, Ken 2004: *Genre and Second Language Writing*. Ann Arbor (MI): The University of Michigan Press.
- Hyltenstam, Kenneth 1984: The use of typological markedness conditions as predictors in second language acquisition: The case of pronominal copies in relative clauses. – Roger Andersen (toim.), *Second Languages: A Cross-Linguistics Perspective* s. 39–58. Rowley (MA): Newbury House.
- Hyon, Sunny 1996: Genre in three traditions: Implications for ESL. – *TESOL Quarterly* 30(4) s. 693–722.
- ICLFI = Kansainvälinen oppijansuomen korpus. Hankkeen verkkosivut.
<http://www oulu.fi/hutk/sutvi/oppijankieli/ICLFI/Yleinen/index.html> (23.11.2011).
- Itkonen, Esa 2005: *Analogy as Structure and Process*. Amsterdam: John Benjamins.
- Itkonen, Esa 2011: Konstruktiokielioppi ja analogia. *Virittäjä* 115 s. 600–605.
- Ivaska, Ilmari 2011: Lausetyyppien sekoittuminen edistyneessä oppijansuomessa – Näkökulmana eksistentiaalilause. – *Lähivertailuja. Lähivörtlusi* 21 s. 65–85.
- Ivaska, Ilmari & Siitonen, Kirsti 2009: Syntaktisesti koodattu oppijankielen korpus: mahdollisuuksia ja ongelmia. – Pille Esilon & Katre Öim (toim.), *Korpusuuringute*

- metodologia ja märgendamise probleemid* s. 54–71. Eesti keele ja kulttuuri instituuti toimetised 11. Tallinn: Tallinna Ülikool.
- Ivaska, Ilmari & Siitonen, Kirsti 2011: Avainrakenneanalyysi: Tapa tutkia oppijankielen lauserakennetta korpusvetoisesti. – *AFinLA-e* 3 s. 35–47.
- Ivaska, Ilmari & Siitonen, Kirsti 2015: Learner language morphology as a window to crosslinguistic similarities. Esitelmä konferenssissa Receptive Multilingualism: Multilingual Resources in Service of Mutual Understanding – REMU 2015. Joensuu, 28.–29.5.2015.
- Jantunen Jarmo Harri 2001: Tärkeä seikka ja keskeinen kysymys: Mitä korpuslingvistinen analyysi paljastaa lähisynonyymeista? – *Virittäjä* 105 s. 170–192.
- Jantunen, Jarmo Harri 2004: Synonymia ja käännössuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaaliin erityispiirteisiin. Joensuun yliopiston humanistisia julkaisuja 35. Joensuu: Joensuun yliopistopaino.
- Jantunen, Jarmo Harri 2007: Oppijansuomen piirteitä korpusvetoisesti. – Pirkko Muikku-Werner, Ossi Kokko & Hannu Remes (toim.), *Virsu III. Suomalais-ugrilaisia kohdekieliä ja kontakteja* s. 69–83. Studies in Languages 42. Joensuu: Joensuun yliopisto.
- Jantunen, Jarmo Harri 2009a: Ei pelkästään mielikuvituksen puutteen vuoksi – Kieliaineistojen systemaattinen käyttö kielentutkimuksessa. – *Virittäjä* 113 s. 101–113.
- Jantunen, Jarmo Harri 2009b: Minulla on aivan paljon rahaa. Fraseologiset yksiköt suomen kielen opetuksessa. – *Virittäjä* 113 s. 356–381.
- Jantunen, Jarmo Harri 2011a: Avainsana-analyysi annotoidun oppijankieliaineiston tutkimisessa: Alustavia havaintoja. – *AFinLA-e* 3 s. 48–61.
- Jantunen, Jarmo Harri 2011b: Kansainvälinen oppijansuomen korpus (ICLFI): typologia, taustamuuttujat ja annotointi. – *Lähivertailuja. Lähivördlusi* 21 s. 86–105.
- Jantunen, Jarmo Harri 2012: Akateemiset ja populaaritekstit korpusvertailussa: sanalistat, avainsanat ja fraseologiset yksiköt. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.), *Genreanalyysi – tekstilajitutkimuksen käytäntöä* s. 186–213. Helsinki: Kotimaisten kielten keskus.
- Jantunen, Jarmo Harri & Brunni, Sisko 2012: Morfologinen priming ja fraseologia vieraan kielen oppimisessa: korpustutkimus oppijansuomesta. – *Lähivertailuja. Lähivördlusi* 22 s. 71–100.
- Jantunen, Jarmo Harri & Piltonen, Saana 2009: Oppijansuomen ja -viron sähköiset tutkimusaineistot. – *Virittäjä* 113 s. 449–458.
- Jantunen, Jarmo Harri & Pirkola, Silja 2015: Oppijansuomen sähköiset tutkimusaineistot. – *Virittäjä* 119 s. 88–103.
- Jarvis, Scott 1998: *Conceptual Transfer in the Interlingual Lexicon*. Bloomington: Indiana University Linguistics Club Publications.
- Jarvis, Scott 2000. Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. – *Language Learning* 50(2) s. 245–309.
- Jarvis, Scott 2010. Comparison-based and detection-based approaches to transfer research. – *EUROSLA Yearbook* 10 s. 169–192.
- Jarvis, Scott 2012. The detection-based approach: An overview. – Scott Jarvis & Scott A. Crossley toim.), *Approaching Language Transfer through Text Classification* s. 1–33. Bristol: Multilingual Matters.
- Jarvis, Scott & Aneta Pavlenko. 2008. *Crosslinguistic Influence in Language and Cognition*. London: Routledge.
- Jensen, Bård Uri 2013: “A chi-square test showed that...” – or did it really? Some remarks on the use of statistical tests in corpus-based research. Plenumsitelmä konferenssissa Learner Corpus Research 2013. Bergen 27.–29.9.2013.

- Jurafsky, Daniel & Martin James R. 2009: *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*.
Toinen, korjattu ja lisätty painos. New Jersey: Pearson Prentice Hall.
- Kaivapalu, Annekatrin 2005: *Lähdekieli kielenoppimisen apuna*. Jyväskylän studies in humanities 44. Jyväskylä: Jyväskylän Yliopisto.
- Kaivapalu Annekatrin & Eslon, Pille 2011: Onko lähisukukielten vaikutus suomen ja viron omaksumiseen symmetristä? Korpuspohjaisen tutkimuksen tuloksia ja haasteita. – *Lähivertailuja. Lähivördlusi* 21 s. 132–153.
- Kajander, Mikko 2013. *Suomen eksistentiaalilause toisen kielen oppimisen polulla*. Jyväskylän Studies in Humanities 220. Jyväskylä: Jyväskylän yliopisto.
- Kalliokoski, Jyrki 1996: Puhe, kirjoitus, tekstilajin normit ja toisella kielellä kirjoittaminen. – Helena Ruuska (toim.), *Moneja baareja: tiellä toimivaan kaksikielisyyteen* s. 107–119. Helsinki: Äidinkielen opettajain liitto.
- Kalliokoski, Jyrki 2006. Tekstilajin taju ja toisella kielellä kirjoittaminen. – Anne Mäntynen, Susanna Shore & Anna Solin (toim.), *Genre - Tekstilaji*. Tietolipas 213. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kalliokoski, Jyrki – Mäntynen, Anne 2004: Tekstilaji ja lähikäsitteitä: määrittelyjä, rajanvetoja. – Hilikka Lamberg, Karl Grönn, Anita Julin, Päivi Kiiski, Mervi Murto, Merja Lumijärvi, Johanna Paalanen & Virpi Ruokola (toim.), *Tekstitaidot puntarissa* s. 8–13. Helsinki: Laatusana Oy.
- Kannisto, Sara 2012: *Substantiivilausekkeiden määritteet edistyneiden suomenoppijoiden kirjoituksessa*. Pro gradu -tutkielma. Turun yliopiston suomen kielen oppiaine.
- Kay, Paul 1995: Construction Grammar. – Jef Verschueren, Jan-Ola Östman & Jan Blommaert (toim.), *Handbook of Pragmatics*. Uudelleenjulkaistu teoksessa *Words and the Grammar of Context* (1997) s. 123–132. Stanford (CA): CSLI Publications.
- Kennedy, Graeme 1991: Between and through: the company they keep and the functions they serve. – Karin Aijmer & Bengt Altenberg (toim.), *English Corpus Linguistics. Studies in Honour of Jan Svartvik* s. 95–110. London: Longman.
- Kenny, Dorothy 2000: Lexical hide-and-seek: Looking for creativity in a parallel corpus. – Maeve Olohan (toim.), *Intercultural Faultlines. Research Models in Translation Studies I. Textual and cognitive aspects* s. 93–104. Manchester: St. Jerome.
- Kortelainen, Kristiina 2011: *Ei ole kaikki muumit laaksossa. Tutkimus suomen kielen idiomikonstruktion produktiivisuudesta*. Pro gradu -tutkielma. Turun yliopiston kieli- ja käännöstieteiden laitos.
- Kotilainen, Lari 2007: *Konstruktoiden dynamiikkaa*. Tohtorinväitöskirja. Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.
- Komppa, Johanna 2012: *Retorisen rakenteen teoria suomi toisena kielenä -ylioppilaskokeen kirjoitelman kokonaisrakenteen ja kappalejaon tarkastelussa*. Tohtorinväitöskirja. Helsingin yliopiston suomen kielen, suomalais-ugrilaisien ja pohjoismaisten kielten ja kirjallisuuksien laitos.
- Kotilainen, Lari 2012: Innovaatio, lipsahdus vai vakiintunut konstruktio? – Ilona Herlin & Lari Kotilainen (toim.), *Verbit ja konstruktiot – tapaustutkimuksia suomesta* s. 201–235. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Laalo, Klaus 2011: *Lapsen varhaiskielioppi ja miniparadigmat*. Suomalaisen Kirjallisuuden Seuran toimituksia 1309. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Lado, Robert 1957: *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.
- LAX = Lauseopin X-arkisto. Kotimaisten kielten tutkimuskeskus - Turun yliopiston kieli- ja käännöstieteiden laitos, Lauseopin arkisto. Turku.
- Lakoff, George 1987: *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: Chicago University Press.

- Langacker, Ronald 1987: *Foundations of Cognitive Grammar*. Vol 1. Stanford: Stanford University Press.
- Langacker, Ronald 1990: Concept, image, and symbol. The cognitive basis of grammar. Berlin: Mouton de Gruyter.
- Langacker, Ronald 2009: Investigations in Cognitive Grammar. *Cognitive Linguistics Research* 42. Mouton de Gruyter, Berliini.
- Larsen-Freeman, Diane 1997: Chaos/complexity and second language acquisition. – *Applied Linguistics* 18(2) s. 141–165.
- Larsen-Freeman, Diane 2002: Making sense of frequency. – *Studies in Second Language Acquisition* 24(2) s. 275–285.
- Larsen-Freeman, Diane 2003. *Teaching Language: from Grammar to Gramming*. Boston: Heinle.
- Larsen-Freeman, Diane 2006: The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. – *Applied Linguistics* 27(4) s. 590–619.
- Larsen-Freeman, Diane & Cameron, Lynn 2008: *Complex Systems and Applied Linguistics*. Oxford: Oxford University Press.
- Latomaa, Sirkku & Tuomela, Veli 1993: Suomi toisena vai vieraana kielenä? – *Virittäjä* 97 s. 238–245.
- Laviosa-Braithwaite, Sara 1996: *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation*. Manchester: UMIST.
- Learner Corpora around the World = Université Catholique de Louvainissa toimivan Centre for English Corpus Linguisticsin ylläpitämä listaus maailman oppijankielen korpuksista. <http://www.uclouvain.be/en-cecl-lcworld.html> (26.12.2014).
- Leino, Pentti 1993: *Suomen kielen kognitiivista kielioppia 1: polysemia – kielen moniselitteisyys*. Kieli 7. Helsinki: Helsingin yliopiston suomen kielen laitos.
- Lieven, Elena, Pine, Julian & Baldwin, Gillian 1997: Lexically-based learning and early grammatical development. – *Journal of Child Language* 24 s. 187–219.
- Lott, David 1983: Analysing and counteracting interference errors. – *English Language Teaching Journal* 37 s. 256–261.
- MacWhinney, Brian 1978: The acquisition of morphophonology. – *Monographs of the Society for Research in Child Development* 43(1).
- MacWhinney, Brian 2001: Emergentist approaches to language. – Joan Bybee & Paul Hopper (toim.), *Frequency and the Emergence of Linguistic Structure* s. 449–470. Amsterdam: John Benjamins.
- Makkonen-Craig, Hannele 2008: Neljä näkökulmaa erääseen tekstilajiin: äidinkielen ylioppilaskokeen tehtävänannot 2003–2006. – *Virittäjä* 112 s. 207–234.
- Marcus, Gary F., Vijayan, S., Bandi Rao, S., & Visthton, P. M. 1999: Rule learning by seven-month-old-infants. – *Science* 283 s. 77–80.
- Marinis, Theodore, Roberts, Leah Felser, Claudia & Clahsen, Harald 2005: Gaps in second language sentence processing. – *Studies in Second Language Acquisition* 27(1) s. 53–78.
- Martin, James R. 1985: Process and text: Two aspects of human semiosis. – James D. Benson & William Greaves (toim.), *Systemic perspectives on discourse. Volume 1: Selective theoretical papers from the 12th International Systemic Workshop* s. 240–270. Norwood (NJ): Ablex.
- Martin, Maisa 1995: *The Map and the Rope. Finnish Nominal Inflection as a Learning Target*. *Studia Philologica Jyväskylän* 38. Jyväskylä: University of Jyväskylä.
- Martin, Maisa, Mustonen, Sanna, Reiman, Nina & Seilonen, Marja 2010: On becoming an independent user. – Inge Bartning, Maisa Martin & Ineke wedder (toim.),

- Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research* s. 57–80. Eurosla: Monograph Series (1).
- Mauranen, Anna 1996. Discourse competence – evidence from thematic development in native and non-native texts. – Eija Ventola & Anna Mauranen (toim.), *Academic Writing: Intercultural and Textual Issues* s. 193–230. Amsterdam: John Benjamins.
- Mauranen, Anna 2000. Strange string in translated language. A study on corpora. – Maeve Olohan (toim.), *Intercultural Faultlines. Research Models in Translation Studies I. Textual and cognitive aspects* s. 119–141. Manchester: St. Jerome.
- McCrostie, James 2008. Writer visibility in EFL learner academic writing: A corpus-based study. – *ICAME Journal* 32 s. 97–114.
- Meunier, Fanny & Granger, Sylviane (toim.) 2008: *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- Meyer, Charles F. 2002: *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University press.
- Miestamo, Matti 2004: Suomen kieltoverbikonstruktio typologisessa valossa. – *Virittäjä* 108 s. 264–388.
- Miller, Carolyn 1984: Genre as social action. – *Quarterly Journal of Speech* 70 s. 151–167.
- Mukherjee, Joybrato & Gries, Stefan Th. 2009: Collostructional nativisation in New Englishes: verb-construction associations in the International Corpus of English. – *English World-Wide* 30(1) s. 27–51.
- Mustonen, Sanna 2015: *Käytössä kehittyvä kieli. Paikat ja tilat suomi toisena kielenä -oppilaiden teksteissä*. Jyväskylä Studies in Humanities 255. Jyväskylä: Jyväskylän yliopisto.
- Määttä, Tuija 2011: Ruotsinkielisten alkeistason suomenoppijoiden paikallissijojen käytöstä. – *Lähivertailuja. Lähivördlusi* 21 s. 154–184.
- Määttä, Tuija 2013: Paikallissijaisten saneiden funktiot ruotsinkielisten alkeistason suomenoppijoiden teksteissä. – *Lähivertailuja. Lähivördlusi* 23 s. 238–264.
- Määttä, Tuija 2014: Kommunikaatioverbeistä ruotsinkielisten suomenoppijoiden teksteissä. – *Lähivertailuja. Lähivördlusi* 24 s. 125–150.
- Nesselhauf, Nadja 2004: *Collocations in a Learner Corpus*. Studies in Corpus Linguistics 14. Philadelphia: John Benjamins.
- Niiranen, Leena 2008: *Effects of learning contexts on knowledge of verbs. Lexical and inflectional knowledge of verbs among pupils learning Finnish in Northern Norway*. Julkaisematon väitöskirja. University of Tromsø. <http://hdl.handle.net/10037/2109> (20.5.2014).
- Odlin, T. 1989: *Language Transfer. Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
- Ojutkangas, Krista 2008: Kielentutkijan mielikuvituksen puute eli mihin todelliseen kielenkäyttöön perustuvaa aineistoa tarvitaan. *Virittäjä* 112 s. 276–284.
- Olshtain, Elite, & Blum-Kulka, Shoshana (1985). Degree of approximation: Nonnative reactions to native speech act behavior. In Susan. M. Gass & Carolyn G. Madden (toim.), *Input in second language acquisition* s. 303–325. Rowley (MA): Newbury House.
- Omar, Alwiya S. 1991: How learners greet in Kiswahili. – Lawrence Bouton & Yamuna Kachru (toim.), *Pragmatics and language learning. Vol 2* s. 59–73. Urbana-Champaign: University of Illinois, Division of English as an International Language.
- Osborne, J. 2013: Comparisons are odorous: native-speaker data in learner corpus research. Plenumsitelmä konferenssissa Learner Corpus Research 2013. Bergen, 27.–29.9.2013.
- Palmer, Frank Robert 1976: *Semantics. A New Outline*. Cambridge: Cambridge University Press.

- Paquot, Magali 2008: Exemplification in learner writing. A cross-linguistic perspective. – Fanny Meunier & Sylviane Granger (toim.), *Phraseology in language learning and teaching* s. 101–119. Amsterdam: John Benjamins.
- Paul, Hermann 1891 [1880]: *Principles of the History of Language*. Englanninkielinen käännös saksankielisen alkuteoksen *Prinzipien der Sprachgeschichte* toisesta painoksesta, kääntänyt H. Strong, W. Logeman & B. Wheeler. London: Longman.
- Penttilä, Rasa 2008. *Suomen ja-konjunktion vastineet ir, o ja pilkku liettuassa*. Pro gradu -tutkielma. Turun yliopiston suomen kielen oppiaine.
- Plakans, Lia & Gebiril, Atta 2013: Using multiple texts in an integrated writing assessment: Source text as a predictor of score. – *Journal of Second Language Writing* 22(3) s. 217–230.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan 1985: *A Comprehensive Grammar of the English Language*. London: Longman.
- Reiman, Nina 2011: Transitiivikonstruktio ikkunana syntaksin kehitykseen: infiniittiset rakenteet ja passiivi taidon indikaattoreina S2-oppijoiden teksteissä. – *AFinLA-e* 3 s. 142–157.
- Reiman, Nina 2014: Yläkoulun S2-oppilaiden transitiivi-ilmausten käyttö Eurooppalaisen viitekehyksen taitotasoilla. – *Lähivertailuja. Lähivördlusi* 24 s. 183–220.
- Ridell, Seija 1994: *Kaikki tiet vievät genreen. Tutkimusretkiä tiedotusopin ja kirjallisuustieteen rajamaastossa*. Tampere: Tampereen yliopiston tiedotusopin laitos.
- Rosch, Eleanor & Mervis, Carolyn B. 1975: Family resemblances: Studies in the internal structure of categories. – *Cognitive Psychology* 7(4) s. 573–605.
- Robinson, Peter 2002: Learning conditions, aptitude complexes and SLA: A framework for research and pedagogy. – Peter Robinson (toim.), *Individual Differences and Instructed Language Learning* s. 113–133. Amsterdam: John Benjamins.
- Saffran Jenny R. 2001: The use of predictive dependencies in language learning, – *Journal of Memory and Language* 44(4) s. 493–515.
- Saffran, Jenny R. & Wilson, Diana P. 2003: From syllabus to syntax: multilevel statistical learning by 12-month-old infants. – *Infancy* 4(2) s. 273–284.
- Salmi, Niina 2010: Virkkeiden ja lauseiden piirteitä edistyneiden suomenoppijoiden teksteissä. Pro gradu -tutkielma. Turun yliopiston suomen kielen oppiaine.
- Saukkonen, Pauli 1984: *Mistä tyyli syntyy?* Helsinki: WSOY.
- Scott, Mike 2010: Problems in investigating keyness, or clearing the undergrowth and marking out trails... – Marina Bondi & Mike Scott (toim.), *Keyness in Texts* s. 43–57. *Studies in Corpus Linguistics* 41. Amsterdam: John Benjamins.
- Scott, Mike & Thompson, Geoff 2001: Introduction: Why ‘patterns of text’? – Mike Scott & Geoff Thompson (toim.), *Patterns of Text. In Honour of Michael Hoey* s. 1–11. Philadelphia: John Benjamins.
- Scott, Mike & Tribble, Christopher 2006: *Textual Patterns: Key Words and Corpus Analysis in Language Education*. *Studies in Corpus Linguistics* 22. Amsterdam: John Benjamins.
- Selinker, Larry 1972: Interlanguage. – *International Review of Applied Linguistics* 10 s. 209–231.
- Seilonen, Marja 2013: *Epäsuora henkilöön viittaaminen oppijansuomessa*. Jyväskylä Studies in Humanities 197. Jyväskylä: Jyväskylän yliopisto.
- Seppälä, Tanja 2013: Oppijansuomen kolligaatit ketjuuntuuissa verbirakenteissa. – *Lähivertailuja. Lähivördlusi* 23 s. 315–340.
- Sharwood Smith, Mike – Kellerman, Eric 1986: Crosslinguistic influence in second language acquisition: an introduction. – Eric Kellerman & Mike Sharwood Smith (toim.), *Cross-linguistic Influence in Second Language Acquisition* s. 1–9. Oxford: Pergamon Press.
- Shore, Susanna & Solin, Anna Solin (toim.) 2006: *Genre – tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura.

- Siitonen, Kirsti & Martin, Maisa 2012: Suomen kielen U-verbit taitotason indikaattorina. – *Lähivertailuja. Lähivördlusi* 22 s. 369–406.
- Siitonen, Kirsti & Niemelä, Jenny 2011: Mitä pitkittäistutkimus voi paljastaa edistyneiden suomenoppijoiden kielitaidosta? – *Lähivertailuja. Lähivördlusi* 21 s. 242–279.
- Siivelt, Keaty & Mustonen, Sanna 2013: Lähdekielen vaikutus ja kielitaitotasot: paikallissijojen kehitys oppijansuomessa. – *Lähivertailuja. Lähivördlusi* 23 s. 341–370.
- Sinclair, John 1966: Beginning the study of lexis. – C. Bazell, J. Catford, M. A. K. Halliday & R. Robins (toim.), *In Memory of J. R. Firth* s. 410–430. London: Longman.
- Sinclair John 1987: Collocation: A progress report. – Ross Steele & Terry Threadgold (toim.), *Essays in Honour of Michael Halliday, vol II* s. 319–331. Amsterdam: John Benjamins.
- Sinclair John 1991: *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair John 1996: EAGLES. Preliminary Recommendations on Corpus Typology. Cambridge Approaches to Linguistics. <http://www.ilc.cnr.it/EAGLES/corpusyp/corpusyp.html> (17.11.2011).
- Sinclair John 2001: Review. – *International Journal of Corpus Linguistics* 6(2) s. 339–359.
- Skehan, Peter 1989: *Individual Differences in Second-language Learning*. London: Edward Arnold.
- Skousen, Royal 1989: *Analogical Modeling of Language*. Dordrecht: Kluwer Academic Publishers.
- Spoelman, Marianne 2010: The copula construction in Finnish learner language: Estonian, German and Dutch learners' use of partitive predicatives. – Pille Eslon & Katrin Õim (toim.), *Korpusuuring ja meetodid* s. 127–152. Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 12. Tallinn: TLÜ Kirjastus.
- Spoelman, Marianne 2011: The use of partitives by Estonian Learners of Finnish at different levels of proficiency. – *Lähivertailuja. Lähivördlusi* 21 s. 280–302.
- Spoelman, Marianne 2013: *Prior Linguistic Knowledge Matters. The Use of the Partitive Case in Finnish Learner Language*. Acta Universitatis Ouluensis B Humaniora 111. Oulu: University of Oulu.
- Stefanowitsch Anatol & Gries, Stefan Th. 2003: Collostructions: investigating the interaction of words and constructions. – *International Journal of Corpus Linguistics* 8(2) s. 209–243.
- Stubbs, Michael 1996: *Text and Corpus Analysis. Corpus-assisted Studies of Language and Culture*. Oxford: Blackwell.
- Stubbs, Michael 2001: On inference theories and code theories: Corpus evidence for semantic schemas. – *Text* 21(3) s. 437–465.
- Stubbs, Michael 2010: Problems in investigating keyness, or clearing the undergrowth and marking out trails. – Marina Bondi & Mike Scott (toim.), *Keyness in Texts* s. 43–58. *Studies in Corpus Linguistics* 41. Amsterdam: John Benjamins.
- Suni, Minna 2012: The impact of Finno-Ugric languages in second language research: Looking back and setting goals. – *Lähivertailuja. Lähivördlusi* 22 s. 407–438.
- Swales, John 1990: *Genre Analysis: English in Academic and Research Setting*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali & Baayen, R. Harald 2012: Models, forests and trees of York English. *Was/were* variation as a case study for statistical practice. – *Language Variation and Change* 24 s. 135–178.
- Tang, Ramona (toim.) 2012: *Academic Writing in a Second or Foreign Language. Issues and Challenges Facing ESL/EFL Academic Writers in Higher Education Context*. London: Continuum International Publishing.
- Taylor, Barry 1975: The use of overgeneralization and transfer learning strategies by elementary and intermediate students of ESL. – *Language Learning* 25(1) s. 73–107.
- Taylor, John R. 2002: *Cognitive Grammar*. Oxford: Oxford University Press.
- Tognini-Bonelli, Elena 1996: *Corpus Theory and Practice*. Birmingham: TWC.

- Tognini-Bonelli, Elena 2001: *Corpus Linguistics at Work*. SCL 6. Benjamins, Amsterdam.
- Tomasello, Michael 1992: *First verbs: a case study of early grammatical development*. Cambridge: Cambridge University Press.
- Tomasello, Michael 1998: The return of constructions. – *Journal of Child Language* 25(2) s. 431–442.
- Tomasello, Michael 2003: *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.
- TOPLING = Toisen kielen oppimisen polut. Tutkimushanke Jyväskylän yliopistossa 2010–2013, hankkeen verkkosivut. <https://www.jyu.fi/hum/laitokset/kielet/topling> (30.8.2012).
- Toropainen, Outi 2014: Pelkoilmaisut S2-oppijoiden taitotasoille arvioiduissa tarinoissa. – *Lähivertailuja. Lähivõrdlusi* 24 s. 244–272.
- Traugott, Elisabeth & Trousdale, Graeme 2013: *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Ukkola, Annette 2009: *Taas yksi nollatutkimus. Substantiivilausekkeiden määritteet S2-oppijoiden teksteissä*. Pro gradu -tutkielma. Jyväskylän yliopiston suomen kielen oppiaine.
- Valtonen, Päivi 2012: *Abiturientti uutistoimittajana. Tekstilajin taju ja uutisen tuottaminen äidinkielen tekstitaidon kokeessa*. Turun yliopiston julkaisuja, C-sarja 350. Turku: Turun yliopisto.
- Vetchinnikova, Svetlana 2014: *Second Language Lexis and the Idiom Principle*. Tohtorinväitöskirja. Helsingin yliopiston nykykielten laitos. Helsinki: University of Helsinki.
- VISK = Auli Hakulinen, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen ja Irja Alho 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkkoversio. <http://scripta.kotus.fi/visk> (17.3.2015).
- Västi, Katja 2012: *Verbittömät Tapahtumanilmaukset. Suunnannäyttäjänä LÄHDE- ja KOHDE-konstruktio*. Acta Universitatis Ouluensis B Humaniora 107. Oulu: University of Oulu.
- Wiersma, Wybo, Nerbonne, John & Lauttamus, Timo 2011: Automatically extracting typical syntactic differences from corpora. – *Literary and Linguistic Computing* 26(1) s. 107–124. Oxford University Press, Oxford.
- Wierzbicka, Anna. 1997: *Understanding cultures through their key words: English, Russian, Polish, German, and Japanese*. New York: Oxford University Press.
- Williams, Raymond 1976: *Keywords. A Vocabulary of Culture and Society*. London: Fontana.
- Wulff, Stefanie, Ellis, Nick. C., Römer, Ute, Bardovi-Harlig, Kathleen & LeBlanc, Chelsea J. 2009: The acquisition of tense–aspect: Converging evidence from corpora and telicity ratings. – *Modern Language Journal* 93(3) s. 336–369.
- Östman, Jan-Ola 2005: Construction discourse: a prolegomenon. – J.-O. Östman & M. Fried (toim.), *Construction Grammars: Cognitive Grounding and Theoretical Extensions*. Amsterdam: John Benjamins.