

# Mahdollisuuden ilmaiseminen S1-suomea ja edistynyttä S2-suomea erottavana piirteenä

ILMARI IVASKA

Turun yliopisto

**Tiivistelmä.** Tässä artikkelissa tarkastellaan edistyneiden suomea toisena kielenään käyttävien (S2) mahdollisuudenilmauksien käyttöä määrällisesti ja verrataan niitä ensikielisten suomenkäyttäjien (S1) mahdollisuudenilmauksiin. Tavoitteena on selvittää, miten mahdollisuutta tyypillisesti ilmaistaan. Erityisesti paneudutaan ilmauksista yleisimpään: *voida*-verbin avulla ilmaistavaan mahdollisuuteen ja sen tyypilliseen käyttöön. Tarkasteltavat tekstit ovat osa Edistyneiden suomenoppijoiden korpusta (LAS2). Ne ovat tenttivastauksia, ja tulokset kuvaavat siis kielenkäyttöä akateemisen kirjoittamisen yhdessä alalajissa.

Mahdollisuutta ilmaistaan S1- ja S2-teksteissä monilta osin samalla tavalla, mutta S2-aineistoissa mahdollisuuden ilmauksia on noin puolet S1-aineiston määrästä. Yleisin mahdollisuudenilmaus on *voida*-verbi, näitä on noin 75 prosenttia kaikista mahdollisuudenilmauksista. Myös *voida*-verbiä käytetään S1- ja S2-teksteissä monilta osin samalla tavalla, mutta S1-aineistossa suurempi osa tapauksista on passiivisia, ja ne kuvaavat jonkinlaista luokittelua ja tiedon hierarkkista jäsentämistä.

Modaalisella verbillä usein ilmaistava kirjoittajan asemoiminen suhteessa sanottuun tietoon on tyypillistä akateemisessa kielessä. Aineistojen välinen ero saattaa johtua kontekstuaalisesta epäkonventionaalisuudesta eli siitä, että S2-kielenkäyttäjät eivät ilmaise yhtä usein tällaista akateemiselle kielelle tyypillistä

kielellistä merkitysvivahdetta, vaikka rakenne on oppijoille sinänsä tuttu.

**Avainsanat:** modaalisuus; suomi toisena kielenä; verbiketju; akateeminen kieli; korpus; kontrastiivinen välikielianaalyysi

## 1. Aluksi

Tarkastelen tässä artikkelissa edistyneen oppijansuomen (S2) mahdollisuuden ilmaisemisen tapoja selvittämällä eri ilmausten tyypillisiä käyttömääriä. Vertaan lisäksi havaintoja ensikieliseen suomeen (S1). Valtaosa tarkastelemistani mahdollisuuden ilmauksista on *voida*-verbin ja *A*-infinitiivin muodostamia verbiketjuja, ja paneudun tutkimuksessani erityisesti tällaisissa rakenteissa esiintyvään vaihteluun ja vaihtelussa havaittaviin aineistojen välisiin eroihin. Esimerkki 1 kuvaa mainittua rakennetta.

(1) Opettaja voi **puhua** helpotettua kieltä --- (S1: las2-vtt01vert050)

Olen valinnut mahdollisuuden ilmaukset tarkasteluni kohteeksi aieman tutkimuksen innoittamana. Olen aiemmassa aineistovetoisessa oppijansuomen tutkimuksessa havainnut *A*-infinitiivien olevan S2:ssa selvästi S1:a harvinaisempia ja todennut eron liittyvän nimenomaan *voida*-verbin sisältämiin mahdollisuutta ilmaiseviin verbiketjuihin (Ivaska 2014a). Tutkimusta motivoi myös se, että modaalisuuden ilmausten käytön ylipäätään on havaittu erottavan eri taitotasoilla olevia S2-oppijoita toisistaan (Kemppainen 2001) ja että lisäksi nimenomaan *voida*-verbin käyttö tietyissä konstruktioissa korreloi ainakin jossain määrin kielitaitotason kanssa (Seilonen 2013). Tarkastelen tässä tutkimuksessa akateemisen kirjoittamisen alalajia: tenttivastauksia. Aiempi tutkimus on osoittanut erityisesti modaalisten verbi-ilmausten olevan tyypillisiä akateemiselle kielenkäytölle (Biber 2006: 15–103). Lisäksi on havaittu, että englantia toisena kielenä kirjoittavat yliopisto-opiskelijat käyttävät vähemmän kirjoittajapositiota ilmaisevia modaali-ilmauksia kuin ensikieliset kirjoittajat (Cotton 2010).

Tavoitteenani on selvittää mahdollisuuden ilmaisemisen tyypillinen variaatio S2:ssa ja verrata sitä S1:een. Tutkimuskysymykseni ovat siis seuraavat:

- 1) Miten paljon ja millä keinoin mahdollisuutta ilmaistaan edistyneiden suomenoppijoiden kirjoitetussa kielessä, ja miten käyttö eroaa ensikielisten teksteistä tai on samankaltaista niiden kanssa?
- 2) Miten mahdollisuuden ilmaisukeinoista yleisintä – *voida*-verbiä – tyypillisesti käytetään tutkimuksen osa-aineistoissa?

Käyttöfrekvenssit ovat keskeisiä keinoja kielen tyypillisen variaation kuvaamisessa. Tarkastellut yksiköt voivat olla niin sanastollisia, rakenteellisia kuin muunkinlaisia kielenpiirteitä. Tällä tavoin oppijankielen ominaispiirteitä ei lähestytä niinkään kohdekielen normien mukaan määriteltyjen laadullisten erityispiirteiden kautta (vrt. Bley-Vroman 1983), eikä tutkimusasetelmasta tule perusluonteeltaan virheanalyttinen. Tällaista tutkimusasetelmaa, jossa korostuu saman kielen useamman varieteetin määrällinen vertailu, kutsutaan oppijankielen korpustutkimuksessa kontrastiiviseksi välikielianaalyysiksi. Vertailtavia varieteetteja voivat olla vaikkapa ensikielinen ja ei-ensikielinen – kuten tässä tutkimuksessa – tai informanttien äidinkielen mukaan erotetut oppijankielen varieteetit. (Granger 1996; 2013.) Tarkastelemalla systemaattisesti kieli-varieteettien sisäisiä jakaumia ja samalla vertailemalla varieteetteja keskenään voidaan ottaa huomioon myös vertailtavissa varieteeteissa (tässä tutkimuksessa S1 ja S2) esiintyvä todellinen variaatio, olipa se luonteeltaan informanttien välistä tai tilanteista (vrt. Osborne 2013). Uskon, että havainnoimalla samanaikaisesti jonkin kielenpiirteen kokonaismäärissä ja sen tyypillisessä käytössä ilmeneviä eroja aineistojen välillä on mahdollista päästä lähemmäs eron taustalla vaikuttavia syitä.

Tutkimusaineistonani on Edistyneiden suomenoppijoiden korpus (LAS2, Ivaska 2014b), johon on kerätty akateemisessa diskurssissa kirjoitettuja suomenkielisiä tekstejä. Käsillä oleva tutkimus on luonteeltaan sikäli korpuspohjainen, että olen rajannut tutkimuskysymykseni ennen aineistoon tutustumista. Kysymykset ovat kuitenkin muotoutuneet

aiemmassa tutkimuksessa tehtyjen määrällisten havaintojen perusteella, mistä syystä alkuperäistä kysymyksenasettelua voidaan pitää korpusvetoisena (termeistä korpuspohjainen ja korpusvetoinen ks. esim. Tognini-Bonelli 2001).

Artikkeli on jäsennelty siten, että luvussa 2 käsittelen tämän tutkimuksen suhdetta kielen teoriaan, mahdollisuuden ilmaisemista osana modaalisuuden kirjoja ja ilmiön aiempia tarkasteluja niin toisen kielen oppimisen kuin S2-tutkimuksenkin kannalta. Luvussa 3 esittelen tarkemmin tämän tutkimuksen aineiston ja ne konkreettiset metodit, joiden avulla pyrin vastaamaan tutkimuskysymyksiin. Luvussa 4 puolestaan raportoin saamani tutkimustulokset ja luvussa 5 teen kokoavaa tarkastelua. Tässä tarkastelussa pohdin yhtäältä sitä, mitä tulokset kertovat S2:n ja S1:n välisistä eroista ja yhtäläisyyksistä, ja toisaalta sitä, mitä ne voivat kertoa toisen kielen oppimisesta ylipäätään.

## 2. Taustaa

### 2.1. Käyttöpohjainen malli ja kompleksisuusteoria

Käyttöpohjaisista malleista puhuttaessa ei varsinaisesti tarkoiteta mitään kaikilta osin yhtenevää koulukuntaa vaan runsaasti näkemyseroja sisältävää löyhää teoreettista viitekehystä, jonka piirissä todellisen käytetyn kielen ja käytössä esiintyneiden ja esiintyvien ilmausten katsotaan luovan kielen rakenteelliset yksiköt. Näin ajatellen kielijärjestelmä koostuu kokoelmasta kielellisiä tapoja, joiden käyttöä ohjaa tyypillisuus ja todennäköisyys. Nämä tavat ja erilaisten tapojen tyypillisuus ja todennäköisyys puolestaan kehkeytyvät (engl. *emerge*) todellisen kielenkäytön pohjalta. (Hopper 1987; Bybee & Thompson 1997.) Kieli siis omaksutaan havainnoimalla käytettyjen ilmausten samankaltaisuutta ja erilaisuutta sekä tekemällä näihin havaintoihin perustuvia yleistyksiä. Tällä tavoin onkin onnistuttu perustellusti kuvaamaan kielellisen syötöksen ja syötöksen laukaiseman omaksumisen välistä suhdetta (esim. Bates & MacWhinney 1987; Tomasello 1992; Lieven ym. 1997). Tämä suhde, joka on kielen omaksumisen keskeisiä mekanismeja, puolestaan perustuu kielen

analogiseen perusuonteeseen eli rakenteellis–funktionaalisiin samankaltaisuuksiin (ks. esim. Skousen 1989; Itkonen 2005).

Kompleksisuusteorialla viitataan laajaan kirjoon kaaos- ja katastrofiteorian kaltaisia ajattelumalleja. Nämä mallit pyrkivät selittämään järjestelmiä, jotka ovat perusuonteeltaan dynaamisia eli joille on ominaista ajassa tapahtuva muutos (yleiskuvan kompleksisista järjestelmistä ja toisen kielen oppimisen suhteesta niihin antaa esim. Larsen-Freeman 1997). Sovellettaessa kompleksisuusteoriaa kielenoppimiseen lähtöoletukset ovat käyttöpohjaisia: oletetaan, että kielijärjestelmä on luonteeltaan avoin ja monimutkainen ja että se syntyy itseohjautuvasti siten, että ilmiöt kehkeytyvät ja mukautuvat dynaamisesti ympäristön ohjaamina. Kieli muotoutuu kontekstissaan sen perusteella, miten sitä käytetään. Kontekstuaalinen omaksuminen puolestaan on mahdollista analogiaan perustuvien mekanismien avulla. Näitä mekanismeja ovat ennen kaikkea imitointi, ennalta tuttujen ja ensi kertaa kohdattujen ilmausten rakenteellisen samankaltaisuuden ja erilaisuuden tunnistaminen, kategorioiden muodostaminen sekä sosiaalisen vuorovaikutustarpeen ohjaama kyky mukautua. (Esim. Larsen-Freeman 1997; Larsen-Freeman & Cameron 2008.) Näiden mekanismien avulla kieleen – tai oikeammin kullekin kielenoppijalle vuorollaan – kehkeytyy rakenteellisia abstraktioita eli konstruktioita, joissa tiettyjen kielellisten ainesten ja tiettyjen merkitysten välille muodostuu pareja (ks. esim. Fillmore 1985; Goldberg 1995; 2006; Croft 2001; Tomasello 2003). Näiden muoto–merkitysparien kompleksisuus voi vaihdella yksittäisistä sanoista (esim. *pallo*) tai morfeemeista (esim. inessiivin päätte *-ssa*) aina hyvin skemaattisiin argumenttirakenteisiin (esim. transitivikonstruktio  $NP_{NOM} + V + NP_{PART/GEN/NOM}$ ). Käyttöpohjaisen teorian mukaan skemaattisetkin konstruktiot kehkeytyvät todellisista kielenilmauksista analogisten mekanismien avulla (ks. esim. Tomasello 1992; Laalo 2011), ja monet osin skemaattisetkin konstruktiot ovat vahvasti sidoksissa tiettyihin leksemeihin (Stefanowitch & Gries 2003; Goldberg 2004).

Korpuksilla tehtävässä oppijankielen tutkimuksessa kielen rakennetta lähestytään usein fraseologiselta kannalta tarkastellen sanojen

yhteisesiintymistä ja siinä esiintyvää vaihtelua, valintarajoituksia ja -preferenssejä (ks. esim. Granger 1998; Nesselhauf 2004; Meunier & Granger 2008). Fraseologisessa näkemyksessä on keskeistä se, että kielen rakenneiden katsotaan usein sijoittuvan leksikaalisesti täysin rajoittamattomien eli vapaiden sanayhdistelmien ja täysin kivettyneiden idiomien väliselle jatkumolle sen mukaan, miten paljon niissä esiintyy vaihtelua (ks. esim. Cowie 1981; Sinclair 1991). Sekä fraseologisen lähestymistavan että konstruktikieliopin näkökulma kieleen onkin tällä tavoin käyttöpohjainen, ja kuten Gries (2008a) huomauttaa, konstruktoiden ja fraseologisten merkitysyksiköiden ero on monilta osin terminologinen: molemmat kuvaavat kielellisen muodon ja merkityksen yhdistymistä, ja rakenne ja sanasto ovat kummassakin toisiinsa kietoutuneita.

Kehkeytyessään konstruktiot voivat stabiloitua omaksumisen etenemisen myötä, mutta vakiintuneetkin muoto–merkitys-parit saattavat muuttua uusien käyttökokemusten myötä (N. Ellis & Larsen-Freeman 2006, ks. myös R. Ellis 1985: 95–96). Konstruktoiden merkitys hahmotuu kontekstuaalisena merkitysyksikkönä (vrt. Jantunen 2004: 29–32; Ivaska 2012), kielellisen lähikontekstin ja laajemman tilannekontekstin kokonaisuutena. Tämän tutkimuksen kannalta on keskeistä, että yhtäältä tutkimusaineiston informantit ovat edistyneitä kielenoppijoita ja että toisaalta aineisto on luonteeltaan melko spesifiä (ks. 3.1). Näin ollen tarkastelussa havaittavat erot S1-aineiston ja S2-aineiston välillä voivat kuvata tilannetta, jossa edistyneiden kielenoppijoiden välikieli ei ole mukautunut uudessa kontekstissa ainakaan täysin ensikielisten kielenkäyttäjien suomen kaltaiseksi, mikä saattaa puolestaan kuvata yleisten kielitaitoerojen lisäksi nimenomaan kontekstin vaikutusta kielitaitoon.

## **2.2. Mahdollisuuden ilmaiseminen osana modaalisuutta**

Ison suomen kieliopin mukaan “modaalisuus on semanttinen alue, jossa on kyse asiointilan todenmukaisuutta ja toteutumismahdollisuuksia koskevista arvioista” (VISK § 1551). Modaalisilla ilmauksilla kielenkäyttäjäsiiis kuvaa omaa näkökulmaansa jostakin asiointilasta, esimerkiksi sen

varmuudesta ja mahdollisuudesta sekä sen luvallisuudesta tai pakollisuudesta. Suomessa modaalisuuden ilmaisemiseen käytettäviä kielellisiä keinoja on runsaasti, ja niiden kirjo ylittää kieliopillistuneista moduksista monenlaisiin leksikaalisiin ilmauksiin (ks. esim. Kangasniemi 1992). Tarkastelen tässä tutkimuksessa edistyneen oppijansuomen mahdollisuuden ilmauksia, joten rajaan tarkastelun mahdollisuutta ilmaiseviin kielen keinoihin. Iso suomen kielioppi (VISK § 1551) listaa seuraavat ilmausjoukot tyypillisiksi tavoiksi ilmaista mahdollisuutta:

**Verbi:** *voida, saada, pystyä, saattaa, mahtaa, ehtiä, tarjeta, viitsiä, tohtia*

**Verbirakenne:** *on ostettavissa ~ saatavilla*

**Adjektiivi:** *mahdollista ~ luultavaa ~ sallittua*

**Modus:** *tulkaa, tulkoon; tullee, tulisi, lienee tullut*

**Adverbi/partikkeli:** *ehkä, varmaan, mahdollisesti, kai*

Edellä oleva mahdollisuuden ilmausten joukko on heterogeeninen: osa ilmauksista on selvästi episteemisiä eli ne arvioivat asiointilan mahdollisuutta tai todennäköisyyttä (Kangasniemi 1992: 147–148; VISK § 1556), toiset ilmaisevat deonttista modaalisuutta eli tahtoa tai normeja (Kangasniemi 1992: 88–89; VISK § 1668), ja jotkut ovat dynaamisia, eli ne ilmaisevat mahdollisuuden suhdetta tilanteen sisäisiin tai ulkoisiin edellytyksiin (Kangasniemi 1992: 19–20; VISK § 1554). Esimerkki 2 kuvaa episteemistä modaalisuutta, esimerkki 3 deonttista modaalisuutta ja esimerkki 4 dynaamista modaalisuutta. Kaikki esimerkit ovat Isosta suomen kieliopista.

(2) Se voi onnistua.

(3) Kunnat saavat itse päättää alkamispäivän.

(4) Lapsi jaksaa juosta pieneen vesialtaaseen koko päivän.

Modaalinen merkitys ilmaisee ylipäätään useiden rinnakkaisten vaihtoehtojen olemassaolon mahdollisuutta (VISK § 1553), ja esimerkiksi tämän tutkimuksen innoittajan ja keskeisen tutkimuskohteen – *voida*-verbin käytön – on todettu eksplisiittisesti olevan modaalisuudeltaan spesifioimatonta niin, että verbiä voidaan käyttää kaikissa modaalisuuden lajeissa (Flint 1980: 4–5; Kangasniemi 1992: 6; VISK § 1566). Lisäksi

kyseessä on melko abstrakti semanttisesti lähimerkityksisten ilmausten joukko, jossa eri ilmausten sisäisen tehtävänjaon voi helposti ajatella kuvastavan oppijankielessä usein havaittuja ilmiöitä, kuten jonkin kielenpiirteen määrällinen tai laadullinen liiakäyttö tai vajaakäyttö (vrt. Levenston 1971). Kangasniemi toteaa lisäksi, että modaalisten ilmausten tulkitseminen mihinkään tiettyyn modaalisuuden lajiin on joskus mahdotonta, sillä ilmaukset saattavat olla monitulkintaisia tai modaalisuuden eri lajit saattavat sulautua merkityksessä yhteen (Kangasniemi (1992: 392–400) käyttää termejä *ambiguity* ja *merger*). Näistä syistä tarkastelen tässä tutkimuksessa mahdollisuuden ilmauksia luokittelematta niitä lähtökohtaisesti eri modaalisuuden lajeihin. Tarkastelen kuitenkin luvun 4 loppuosassa esimerkinomaisesti kahta ilmausta ja pohdin myös niiden suhdetta yllä kuvattuun modaalisuuden eri lajien jakoon.

### **2.3. Modaalisuus akateemisessa kirjoittamisessa ja toisen kielen oppimisen tutkimuksessa**

Tässä tutkimuksessa tarkasteltavan tekstilajin – akateemisen kirjoittamisen ja sen alalajin eli tenttivastauksen – kannalta mahdollisuuden ilmaiseminen on keskeistä. Rinnakkaisten mahdollisuuksien ilmaiseminen on tyypillistä akateemiselle kirjoittamiselle, jonka keskeisiä elementtejä ovat yhtäältä aiemman tutkimuksen tuntemus ja esittely ja toisaalta kirjoittajan itsensä asemoiminen ja etäännyttäminen tästä aiemmin sanotusta (engl. *stance*, ks. esim. Campbell 1975; Hyland 1999). Hyland (mt. 101) katsoo kirjoittajaposition koostuvan kolmesta osasta: todistavuudesta (engl. *evidentiality*), asenteesta (engl. *affect*) ja suhtautumisesta (engl. *relation*). Esimerkiksi akateemisen englannin osalta on osoitettu, että nimenmaan modaaliset verbit ovat selvästi tyypillisin tapa ilmaista kielenkäyttäjän omaa asemaa riippumatta siitä, mistä akateemisen kielenkäytön rekisteristä on kyse (Biber 2006: 95–103).

Suomen kielen osalta en tunne yhtään akateemista kieltä täsmälleen tältä kannalta tarkastelevaa tutkimusta. Jantunen vertaa kirjoitettua akateemista suomea populaariteksteihin ja tutkii tyypillisimpiä sanastollisia



eroja näiden välillä. Yksittäisten sananmuotojen frekvenssien tarkastelussa käy ilmi, että *voi* on tyypillisesti yleisempi populaariteksteissä mutta *voidaan* taas akateemisissa teksteissä. (Jantunen 2012: 202–203.) Mauranen tutkii akateemisen kirjoittamisen tyypillisiä retorisia keinoja ensikielenään englantia käyttävien englanninkielisissä teksteissä, ensikielenään suomea käyttävien suomenkielissä teksteissä ja ensikielenään suomea käyttävien englanninkielisissä teksteissä. Mauranen nostaa tyypillisiksi keinoiksi selityksen (engl. *explanation*), päättelyn (engl. *conclusion*), tulosten raportoinnin (engl. *reporting current finding*), aiemman tutkimuksen raportoinnin (engl. *reporting previous research*), yleisesti tunnettujen seikkojen raportoinnin (engl. *reporting general facts or principles*), metodien raportoinnin (engl. *reporting methods*) ja arvioinnin (engl. *evaluation*). Mauranen ei erikseen nosta modaalisuutta esiin, mutta edellä mainittua seitsemää kategoriaa kuvaavista esimerkeistä neljässä käytetään modaalisia keinoja osana ilmausta. (Mauranen 1993: 234–235.) Saukkonen (1984: 61–62) puolestaan mainitsee abstraktiuden ja ilmaisurakenteen monimutkaisuuden olevan ominaista tieteellis-teoreettiselle tyyliäjille. Kuvaavaa on, että Saukkosen seitsemästä lauseesta koostuvassa tieteellis-teoreettisen tyylin esimerkkikatkelmassa kolmessa lauseessa on finiittisenä verbinä modaalinen *voida* (mt. 35–36).

Modaalisuuden ilmaukset voidaan siis nähdä akateemiselle kirjoittamiselle tyypillisenä piirteenä. Toisaalta Mauranen osoittaa kuitenkin tekstilajikohtaisten tekstuaalisten piirteiden olevan ennen kaikkea kieli-taitokysymys. Hän vertaa ensikielisten kirjoittajien englanninkielisiä ja suomenkielisiä akateemisia tekstejä toisiinsa tekstuaalisen teemarakenteen osalta ja vertaa niitä lisäksi ensikieleltään suomalaisten englanniksi kirjoittamiin akateemisiin teksteihin. Mauranen päätyy toteamaan, että ei-ensikielisten kirjoittamat tekstit poikkeavat teemarakenteiltaan ensikielisten kirjoittamista ja että ensikielisten tekstien välillä, kielestä riippumatta, tällaista eroa ei ole. (Mauranen 1996.) Edistyneen oppijanenglannin osalta Cotton (2010) puolestaan osoittaa, että Australiassa opiskelevat, ensikielenään jotain muuta kuin englantia

puhuvat jatko-opiskelijat<sup>1</sup> käyttävät kurssitehtävissään modaalisia verbejä selvästi ensikielisiä jatko-opiskelijoita vähemmän.

S2-tutkimuksessa modaalisuuden ilmausten oppimista ja käyttöä ylipäätään on tutkittu tai sitä on sivuttu aiemminkin. Esimerkiksi Kemppainen (2001) toteaa eri kielitaitotasolla olevien suomenoppijoiden ilmaisevan modaalisuutta eri tavoin siten, että modaalisuuden ilmaiseminen yleistyy kirjoitetussa kielessä alkeistason jälkeen ja että edistyneillä oppijoilla modaalisuuden ilmaisukeinot ovat monilta osin ensikielisten kielenkäyttäjien ilmaisukeinojen kaltaisia. Niiranen (2008: 300–302) on puolestaan havainnut oppimiskontekstin vaikuttavan suomen modaalisten verbikonstruktioiden käyttöön siten, että kaksikielisessä ympäristössä suomea oppineet käyttävät modaalisesta verbistä ja A-infinitiivistä koostuvaa verbikonstruktiota selvästi enemmän kuin muodollisessa luokkahuoneympäristössä suomea opiskelleet.

Seppälä (2012: 41–42, 84–85) tarkastelee erilaisia kirjoitetun S2-kielen ketjuuntuvia verbirakenteita rektioiden kannalta ja havaitsee, että *voida*-verbin käyttö on verbirakenteissa laadullisesti S1-kielen kaltaista. Seilonen (2013) puolestaan toteaa yleiseurooppalaista kielitaitotasojen viitekehystä (CEFR) soveltavassa, oppijansuomen epäsuoraa henkilöarviointia käsittelevässä tutkimuksessaan, että nollapersoonainen, tyypillisesti *voida*-verbin yksikön kolmannen persoonan muodosta ja A-infinitiivistä koostuva *voi*-konstruktio yleistyy aikuisten suomenoppijoiden kirjoituksessa B2-tasolle asti mutta että se hallitaan laadullisesti jo A1-tasolta alkaen (mt. 103–107). Passiivisia *voida*-verbejä puolestaan aletaan ylipäätään käyttää vasta B1-tasolla ja muita passiivisia modaali-verbejä vasta C1-tasolla (mt. 64–69).

Aiempien tutkimusten aineistot poikkeavat luonteeltaan akateemisesta kielestä: Niirasan aineisto on kerätty 12–16-vuotiailta nuorilta, kun taas Seilosen aikuisia koskeva aineisto ja Kemppaisen koko aineisto ovat Yleisten kielitutkintojen suomen kirjoittamisen osakokeen suorituksia eri testitasoilta. Seppälän aineisto puolestaan on osa

---

<sup>1</sup> Engl. *graduate students*, voinee tässä tapauksessa tarkoittaa sekä maisteri- että tohtoritason opiskelijoita.

Kansainvälistä oppijansuomen korpusta (ICLFI, ks. Jantunen 2011), jonka informantit opiskelevat suomea vieraana kielenä ulkomaisissa yliopistoissa. Seppälän tutkimuksessa ei myöskään ole eroteltu kirjoittajia kielitaitotasoin. Aiemmat tutkimukset kuvannevat kuitenkin sitä, milloin keinot mahdollisuuden ilmaisemiseen ylipäättään kehkeytyvät, ja sitä, että mahdollisuuden ilmausten on havaittu erottavan S1- ja S2-kielenkäyttäjiä tai eritasoisia S2-kielenkäyttäjiä toisistaan. Lisäksi voidaan olettaa, että modaalisuuden ilmaiseminen ja erilaiset *voida*-rakenteet ovat edistyneille S2-kielenkäyttäjille ainakin laadullisesti tuttuja.

### 3. Aineisto ja metodit

#### 3.1. Aineistona edistyneiden suomenoppijoiden korpus

Tämän tutkimuksen aineisto on osa Turun yliopiston Edistyneiden suomenoppijoiden korpusta (LAS2, ks. tarkemmin Ivaska 2014b). Toistaiseksi kaikki korpuksen tekstit on kerätty sellaisilta ei-ensikielisiltä suomenkäyttäjiltä, jotka ovat aineistoa kerätessä suorittaneet suomalaisessa yliopistossa maisterintutkintoa suomeksi. Kaikki korpuksen tämänhetkiset informantit ovat opiskelleet suomea sekä Suomessa että Suomen ulkopuolella. Valtaosa informanteista on suorittanut alemman korkeakoulututkinnon jossain muualla kuin Suomessa. Korpus on luonteeltaan moniensikielinen. Lisäksi korpukseen kuuluu suomea ensikielenään käyttäviltä opiskelijoilta kerätty vertailuaineisto. Toistaiseksi valtaosa aineistosta käsittelee aihepiiriltään joko kielitiedettä tai muita humanistisia aihepiirejä. Kaikki tekstit on alun perin kirjoitettu osana informanttien korkeakouluopintoja, ja niiden ensisijainen funktio ei ole ollut kielitaidon osoittaminen. Kaikilta korpukseen aineistoa luovuttaneilta S2-informanteilta on arvioitu kaksi tekstiä eurooppalaisen kielitaidon viitekehyksen (CEFR) mukaisesti. Arviot jakautuvat tason B1 ja C2 välille niin, että valtaosa teksteistä on arvioitu joko B2- tai C1-tasoisiksi (arvioiden jakauma 19.12.2013: B1 5 %, B2 35 %, C1 55 %, C2 6 %). Tämän perusteella aineistoa voitaneen kutsua taitotasoltaan edistyneeksi.

LAS2 sisältää useita tekstilajeja, mutta tämän tutkimuksen koko aineisto on tenttivastauksia. Aineistossa on yhteensä 436 tekstiä. Taulukko 1 kuvaa tarkemmin aineiston kokoa ja sen jakautumista tutkimus- ja vertailuaineistoon.

**TAULUKKO 1.** *Tässä tutkimuksessa käytettävä aineisto*

Osa-aineisto	Tekstiyksiköt	Ensikielet	Saneet
S2-aineisto	334	tšekki, englanti, islanti, japani, komi, liettua, puola, saksa, slovakki, unkari, viro, venäjä	145 379
S1-aineisto	102	–	58 649
<b>Yhteensä</b>	<b>436</b>	–	<b>204 028</b>

Aineisto on kokonaisuudessaan lemmattu, ja se on annotoitu sanaluokkien, sanojen morfologisten muotojen ja syntaktisten funktioiden osalta. Aineisto on lisäksi rakenteistettu siten, että se on jaoteltu kappaleisiin, virkkeisiin ja lauseisiin. Lisäksi tekstien otsikot, kokonaiset esimerkkilauseet, taulukot ja ranskalaisten viivojen avulla muotoillut listat yms. on merkitty siten, että aineistoon tehtävät haut eivät kohdistu niihin. Aineisto on tallennettu XML-muotoon, ja se seuraa muiden Turun yliopistossa koostettujen korpusten rakenteellisia ratkaisuja (vrt. Lauseopin X-arkisto: LaX; Mikael Agricolan morfosyntaktinen tietokanta: Inaba 2007). Kaikki aineistohaut on tehty LAS2:n omilla hakutyökaluilla (hakutyökaluista tarkemmin ks. Ivaska 2014b) ja kaikki tilastolliset analyysit on tehty R-ohjelmistoympäristöä käyttäen (R 2013).

### 3.2. Kielenpiirteen kontekstuaalisen profiilin tarkastelu

Keräsin aineistosta kaikki sellaiset mahdollisuuden ilmaukset, jotka sisältyvät Isossa suomen kieliopissa (VISK § 1551) mainittuihin mahdollisuuden ilmauksen keinoihin suomen kielessä. Pysin analyysissä määrälliseen aineistolähtöisyyteen, enkä ole systemaattisesti läpikäynyt jokaista havaintoa. Tästä syystä olen jättänyt tarkastelun ulkopuolelle VISK:ssa mahdollisuutta kuvaavina mainitut verbi-ilmaukset *saada* ja

*ehsiä* sekä moduksista imperatiivin ja konditionaalin, sillä niillä on runsaasti myös muuta kuin mahdollisuutta ilmaisevaa käyttöä. Lopulliset hakuni olivat seuraavat<sup>2</sup>:

**Leksikaaliset ilmaukset:** verbit *voida, pystyä, saattaa, mahtaa, tarjeta, viitsiä* ja *tohtia*; adjektiivit *mahdollinen, luultava* ja *sallittu*; adverbit *ehkä, varmaan, mahdollisesti* ja *kai*.

**Moniosaiset rakenteet:** *olla* V-PASS-PCP1-INE/ADE (*on ostettavissa, on saatavilla*).

**Modukset:** potentiaali.

Tarkastelen hakutulosten perusteella modaalisuuden ilmaisemisen määrällisiä eroja S2:n ja S1:n välillä. Tutkimus seuraa perusasetelmaltaan niin sanottua kontrastiivista välikielianalyysia (engl. *contrastive interlanguage analysis*, CIA, Granger 1996; 2013). Metodina tavoitteena on verrata saman kielen kahta kahta eri varieteettia. Kuten Granger huomauttaa, riittävän suuret korpusaineistot mahdollistavat yli- ja aliedustusten kvantitatiivisen tarkastelun (Granger 1996: 45) ja auttavat erottamaan tarkasteltavalle kielenpiirteelle ominaisen vaihtelun varieteettisidonnaisesta vaihtelusta (Granger 2013).

Valtaosa aineiston tekstiyksiköistä on 500–1000 saneen mittaisia. Kaikki frekvenssihavainnot on normalisoitu 1000:ta sanaa kohti kussakin tekstissä (havaintojen määrä / tekstin pituus × 1000), ja aineistojen välisten frekvenssien vertailu tehdään siis näiden normalisoitujen frekvenssien pohjalta. Näin voidaan välttää yksittäisten tekstien liian suuri vaikutus (vrt. Gries 2008b). Tässä tutkimuksessa tekemäni havainnot ovat monilta osin luonteeltaan kuvailevia, mutta tilastollisia päätelmiä tehdessäni olen soveltanut Mann-Whitneyn *U*-testiä. Olen valinnut

<sup>2</sup> Hakuteknisistä syistä kaikissa verbin sisältävissä mahdollisuuden ilmauksissa verbi määriteltiin lauseen syntaktiseksi predikaatiksi. Tästä syystä analyysini käsittelee vain affirmatiivisia lauseita. Selvitin kuitenkin yleisimmän mahdollisuuden ilmauksen, *voida*-verbin, kokonaisjakauman negatiivisessa kontekstissa, eikä se eroa kaikkien mahdollisuuden ilmausten jakaumasta (ks. alaviite 4). Symmetrian vuoksi sisällytin kaikki kriteerit täyttävät ilmaukset mukaan hakuheitoihin. Kaikkia haettuja tapauksia (kuten *tarjeta*- ja *tohtia*-verbejä) ei kuitenkaan löydy aineistosta.

testin, sillä se on luonteeltaan epäparametrinen eikä näin ollen aseta havaintojen frekvenssijakaumalle rajoitteita (ks. Hollander & Wolfe 1972: 68–75). Arvioidessani havaintojen tilastollista merkitsevyyttä pidän merkitsevyyden raja-arvona  $p < 0.05$ .

Tarkastelen seuraavassa lähemmin mahdollisuudenilmauksista yleisimmän – *voida*-verbin – kontekstuaalista profilia. Jantusen käyttämässä kuvauksessa leksikaalinen merkitysyksikkö koostuu jonkin leksikaalisen yksikön sisäisestä vaihtelusta, sen tyypillisestä kielellisestä esiintymisympäristöstä eli kotekstista ja laajemmasta tilannekontekstista (Jantunen 2004: 31). Tilannekonteksti on käyttämässäni aineistossa ver-raten homogeeninen, joten keskityin analyysissä ennen kaikkea sisäisen ja ulkoisen vaihtelun tarkastelemiseen. Seuraan pitkälti Francisin (1993) esittämää korpusvetoista tapaa kuvata jonkin kielenpiirteen tyypillistä käyttöä tarkastelemalla siinä esiintyvää sisäistä ja ympäristössä havaittavaa vaihtelua (engl. *item and environment*). Käytännössä tarkastelen sitä, missä morfologisissa muodoissa *voida*-verbi esiintyy, mitkä ovat verbi-rakenteen tyypillisimpiä semanttisia pääverbejä ja millaisessa leksikaalis–morfologis–syntaktisessa kotekstissa *voida*-verbi yleensä esiintyy<sup>3</sup>.

## 4. Tulokset

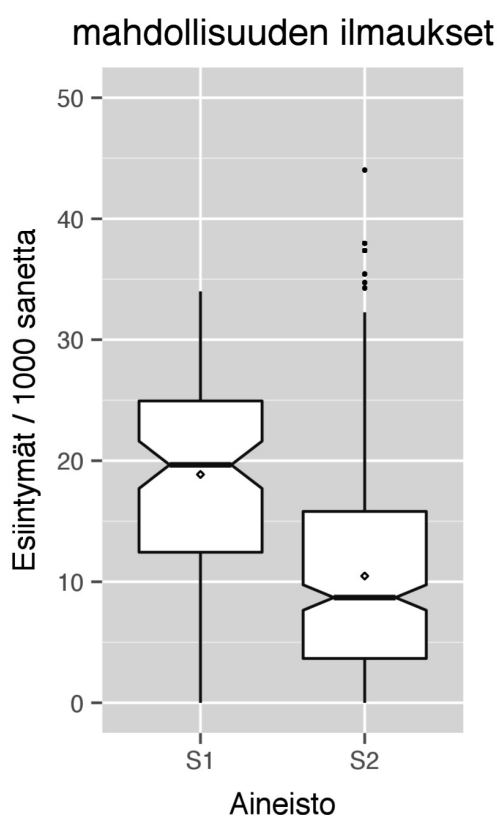
### 4.1. Mahdollisuuden ilmaukset aineistossa

Aineistossa on yhteensä 2 668 edellä kuvatut kriteerit täyttävää mahdollisuuden ilmausta (13,1 / 1 000 sanetta). S1-aineistossa (18,9 / 1 000 sanetta) näitä on lähes kaksi kertaa niin paljon kuin S2-aineistossa (10,6 / 1 000 sanetta), ja aineistojen välinen ero on tilastollisesti hyvin merkitsevä ( $U = 26226$ ,  $p < 2.2e^{-16}$ ,  $r = .40$ )<sup>4</sup>. Kuvio 1 havainnollistaa frekvenssien jakautumista S1- ja S2-aineistoissa.

---

<sup>3</sup> Kotekstista puhuessani käytän termejä L2, L1, R1 ja R2. Nämä viittaavat sana-asemiin tarkasteltavan sanan vasemmalla (L1 ja L2) ja oikealla (R1 ja R2) puolella.

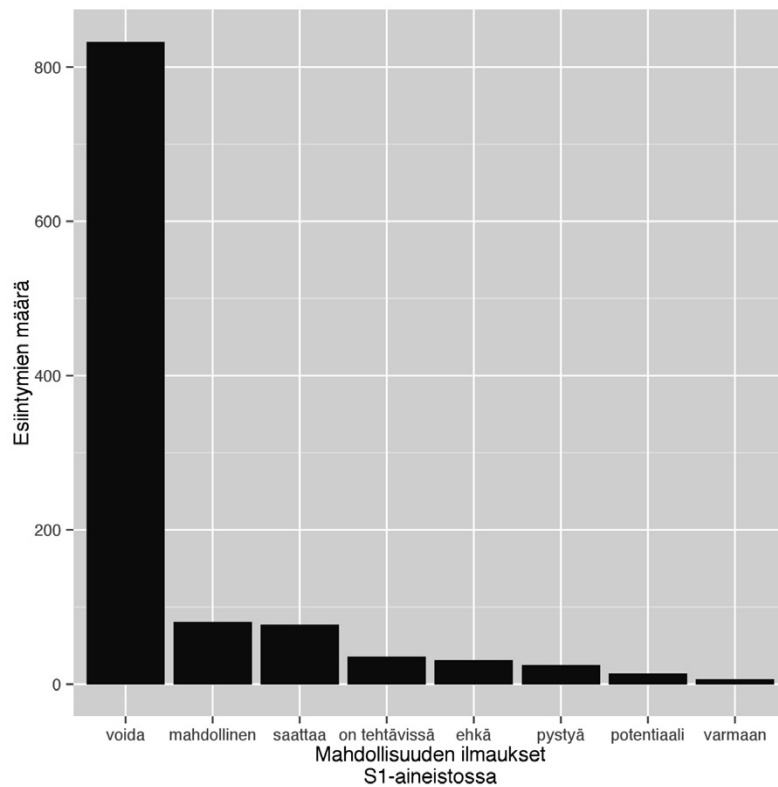
<sup>4</sup> Tarkastelin tässä vaiheessa myös negatiivisten *voida*-ilmausten määrää. Jakauma ei poikennut mahdollisuuden ilmausten kokonaiskuvasta (S1: 1,1 / 1 000 sanetta; S2: 0,6 / 1 000 sanetta).



**KUVIO 1.** Mahdollisuuden ilmausten määrä tutkimuksen osa-aineistoissa<sup>5</sup>

Noin kolme neljäsosaa kaikista mahdollisuuden ilmauksista on *voida*-verbejä, eikä S1- ja S2-aineistojen välillä ole juuri eroa. Eri mahdollisuuden ilmausten jakauma on aineistojen kesken hyvin samankaltainen. Tutkimuksen lähtökohtana ollut havainto, että *voida*-verbistä ja infinitiivistä koostuvaa rakennetta esiintyy S2-aineistossa vähemmän, ei siis selity sillä, että jotain muuta mahdollisuuden ilmausta käytettäisiin sen sijasta. Kuviot 2 ja 3 kuvaavat yleisimpien mahdollisuuden ilmausten jakaumaa kummassakin osa-aineistossa. Näiden havaintojen valossa näyttää siltä, että mahdollisuuden ilmaukset ovat kokonaisuudessaan harvinaisempia S2-aineistossa.

<sup>5</sup> Laatikkojanakuvaajissa valkoinen osa kuvaa vaihtelun keskimmäistä 50:tä prosenttia ja laatikosta lähtevä viiva sen ulkopuolista vaihtelua. Voimakkaasti poikkeavat havainnot on merkitty erillisillä pisteillä.



**KUVIO 2.** Yleisimmät mahdollisuuden ilmaukset S1-aineistossa

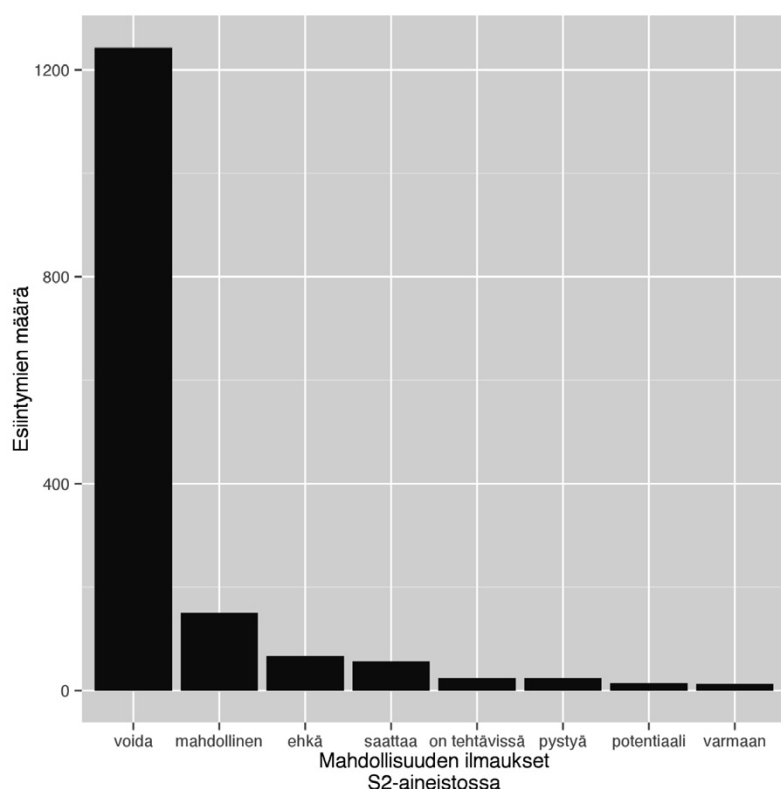
#### 4.2. *voida*-verbin modaalinen käyttö

##### S1- ja S2-suomea erottavana kielenpiirteenä

Aineiston kaikista mahdollisuuden ilmauksista likimain kolme neljäsosaa (yht. 2076 tapausta; 9,9 / 1 000 sanetta) koostuu *voida*-verbin sisältävistä verbiketjuista. Kuten aiempi tutkimus antaa olettaa (Ivaska 2014a), S1- ja S2-aineiston välinen ero vastaa mahdollisuuden ilmauksien yhteismäärien eroa, sillä tällaiset verbiketjut ovat S1-aineistossa S2-aineistoa yleisempiä (S1: 14,7 / 1 000 sanetta; S2: 8,4 / 1 000 sanetta) ja ero on tilastollisesti hyvin merkitsevä ( $U = 24980$ ,  $p = 8.79e^{-13}$ ,  $r = .34$ ).

*voida*-verbiin sisäisen morfologisen vaihtelun tarkastelu osoittaa, että verbi on yleensä joko indikatiivin yksikön kolmannen persoonan preesensissä (<fin ind pres sg3>) tai indikatiivin passiivin preesensissä (<fin pass ind pres>). Taulukko 2 kuvaa sisäistä jakaumaa tarkemmin.





**KUVIO 3.** Yleisimmät mahdollisuuden ilmaukset S2-aineistossa

**TAULUKKO 2.** *voida*-verbin yleisimmät morfologiset muodot *voida*-verbistä ja toisesta verbistä koostuvissa verbiketjuissa

<i>voida</i> -verbin morfologinen muoto	S1-aineisto	S2-aineisto
<fin ind pres sg3>	49 %	63 %
<fin pass ind pres>	27 %	15 %
<fin ind pres pl3>	12 %	13 %
<fin cond pres sg3>	7 %	4 %
<fin pass cond pres>	2 %	< 1 %
[---]	[---]	[---]
<b>Yhteensä</b>	<b>100 %</b>	<b>100 %</b>

Kuten jakaumat osoittavat, verbien sisäinen vaihtelu on tarkastelluissa aineistoissa samankaltaista, mutta passiivimuotoiset *voida*-verbit ovat yleisempiä S1-aineistossa. Esimerkit 5 ja 6 havainnollistavat passiivissa olevaa tapausta.

- (5) Sääverbejä **voidaan käyttää** myös metaforissa tai polysemisesti, jolloin subjektina voi toimia jokin muukin kuin varsinaisesti säähän liittyvä tarkoite. (S1: las2-vtt01vert116)
- (6) Suomessa on b, d ja g, jotka ovat klusiilien heikkoastevaihtelun pareina, sekä h, jota **voidaan sanoa** joko soinnilliseksi tai soinnittomaksi. (S2: las2-12tt01te05)

Passiivin kannalta on mielenkiintoista, että – kuten aiemmassa tutkimuksessa on todettu – passiivimuotoisten verbien käyttö S2-oppijoiden kirjoituksissa alkaa vasta B-tasolla (Seilonen 2013: 64–69) ja niiden kokonaismäärä kasvaa aina C-tasolle asti (mt. 58–59). Tämän tutkimuksen aineistossa passiivin käytössä ei kokonaisuudessaan ole kovin suurta eroa aineistojen välillä (S1: 17,5 / 1 000 sanetta; S2: 15,2 / 1 000 sanetta), ja tämä ero selittyy nimenomaan *voida*-verbin passiivimuotojen frekvenssierolla (S1: 4,2 / 1 000 sanetta; S2: 1,4 / 1 000 sanetta). Jantusen mukaan *voidaan*-sanamuoto on suomessa tyypillinen nimenomaan akateemisissa teksteissä (Jantunen 2012: 202–203), ja S1- ja S2-aineiston välinen ero voikin tämän valossa liittyä nimenomaan tekstilajille tyypilliseen kielenkäyttöön.

Tarkastelin seuraavaksi, minkälaisien verbien yhteydessä modaaliset *voida*-ilmaukset yleensä esiintyvät. Kuten taulukon 3 jakaumat osoittavat, molemmissa aineistoissa yleisin pääverbi on *olla*-verbi.

**TAULUKKO 3.** Yleisimmät verbit *voida*-verbistä ja toisesta verbistä koostuvissa verbiketjuissa

S1-aineisto	S2-aineisto
<i>olla</i> 25 %	<i>olla</i> 30 %
<i>käyttää</i> 4 %	<i>sanoa</i> 4 %
<i>luokitella</i> 4 %	<i>esiintyä</i> 3 %
<i>jakaa</i> 3 %	<i>käyttää</i> 3 %
<i>saada</i> 2 %	<i>nähdä</i> 2 %
[---]	[---]
<b>100 %</b>	<b>100 %</b>

On huomionarvoista, että S1-aineistossa yleisimpien verbien joukossa on useampia jonkinlaista luokittelua kuvaavia verbejä, kuten *luokitella* ja *jakaa* sekä myös *erottaa*, *muodostaa* ja *jaotella*. S2-aineistosta tämä ilmausryhmä ei näytä nousevan esiin. Esimerkit 7 ja 8 kuvaavat tällaisia ilmauksia.

- (7) Ikoni, indeksi ja symboli ovat keskeisiä kielen käsitteitä, joiden avulla sanoja voi **luokitella**. (S1: las2-vtt01vert107)
- (8) Ennen kaikkea sanat voidaan **jakaa** niiden taivutusmuotojen perusteella. (S2: las2-14tt01te05)

Myös *voida*-verbien tyypilliset esiintymiskontekstit kuvastavat sitä, että käyttöero liittyy ennen kaikkea passiivissa esiintyviin ilmauksiin. *voida*-verbien tyypillisen kielellisen lähikontekstin eli kotekstin tarkastelu syntaktisten funktioiden osalta (ks. taulukot 4 ja 5) osoittaa, että koteksti on S1- ja S2-aineistoissa hyvin samankaltainen: verbiä edeltää useimmiten nominaalinen subjekti (<npsubj>, *kielitaito* esimerkissä 9, *internet* esimerkissä 10), jota puolestaan edeltää tyypillisesti joko lauseenalku (<cl>, kuten esimerkissä 10) tai substantiivinmäärite (<nmod>, *lääkäreiden* esimerkissä 9). *voida*-verbiä seuraa tyypillisesti predikaatin toinen osa (<pred2>, *vaihdella* esimerkissä 9, *antaa* esimerkissä 10), jota puolestaan seuraa useimmiten joko adverbiaali (<advl>, *paljonkin* esimerkissä 9) tai substantiivinmäärite (*hyvän* esimerkissä 10).

- (9) Esimerkiksi lääkäreiden kielitaito voi vaihdella paljonkin --- (S1: las2-vtt01vert073)
- (10) Internet voisi antaa hyvän pohjan omankielisille keskusteluille. (S2: las2-33tt01te07)

**TAULUKKO 4.** *voida-verbin tyypillinen kolligationaalinen konteksti S1-aineistossa syntaktisten funktioiden osalta*

L2-asema S1-aineistossa	L1-asema S1-aineistossa	<i>voida-</i> verbi	R1-asema S1-aineistossa	R2-asema S1-aineistossa
<cl> 33 %	<npsubj> 36 %		<pred2> 78 %	<advl> 27 %
<nmod> 29 %	<npobj> 24 %		<advl> 13 %	<nmod> 21 %
<advl> 14 %	<advl> 21 %		<nmod> 2 %	<pred2> 12 %
<lause-cnj> 7 %	<nmod> 5 %		<infsubj> 1 %	<npobj> 7 %
<s> 4 %	<cl> 4 %		<npobj> 1 %	</cl> 7 %
[---]	[---]		[---]	[---]

**TAULUKKO 5.** *voida-verbin tyypillinen kolligationaalinen konteksti S2-aineistossa syntaktisten funktioiden osalta*

L2-asema S2-aineistossa	L1-asema S2-aineistossa	<i>voida-</i> verbi	R1-asema S2-aineistossa	R2-asema S2-aineistossa
<cl> 33 %	<npsubj> 42 %		<pred2> 84 %	<advl> 22 %
<nmod> 29 %	<advl> 26 %		<advl> 9 %	<nmod> 22 %
<advl> 10 %	<npobj> 12 %		<infsubj> 2 %	</cl> 13 %
<lause-cnj> 8 %	<cl> 7 %		<pred> 1 %	<npobj> 9 %
<s> 7 %	<advl:p> 5 %		<nmod> 1 %	<pred2> 8 %
[---]	[---]		[---]	[---]

Kuten taulukoista 4 ja 5 näkyy, *voida*-verbiä edeltävässä L1-asemassa esiintyy subjektin lisäksi melko usein myös adverbiaali tai nominaalinen objekti (<npobj>): S1-aineistossa L1-asemassa on nominaalinen objekti 24 %:ssa tapauksista, kun S2-aineistossa näin on vain 12 %:ssa tapauksista<sup>6</sup>. Tapaukset ovat tyypillisesti passiivissa olevia lauseita,

<sup>6</sup> Tein vastaavan kontekstuaalisen tarkastelun kaikilla annotoinnin tasoilla eli niin lemموjen, sanaluokkien morfologisten muotojen kuin syntaktisten funktioidenkin osalta. Tässä syntaktisten funktioiden osalta kuvatut tendenssit heijastuvat myös muista annotoinnin tasoista, mutta syntaktiset funktiot kuvaavat eroa parhaiten. Tästä syystä raportoin tässä kontekstuaalisen profiilin ainoastaan tältä osin.

joissa objekti esiintyy lauseen teemapaikalla (teemapaikasta ks. VISK § 1370), kuten edellä olleissa esimerkeissä 5, 6 ja 8. On lisäksi mielenkiintoista, että edellä kuvatut luokittelua ilmaisevat verbit esiintyvät tyyppillisesti nimenomaan verbiketjuissa, joissa *voida*-verbi on passiivissa, ja että nämä tapaukset kattavat S1-aineiston passiivisista verbiketjuista huomattavasti suuremman osan kuin S2-aineiston passiivisista verbiketjuista. Taulukko 6 kuvaa yleisimpiä tällaisten verbiketjujen frekvenssejä aineistoissa.

**TAULUKKO 6.** 10 yleisintä verbiä voidaan-verbimuodosta ja toisesta verbistä koostuvissa verbiketjuissa

S1-aineisto	S2-aineisto
<i>jakaa</i> 10 %	<i>sanoa</i> 11 %
<i>luokitella</i> 8 %	<i>jakaa</i> 7 %
<i>käyttää</i> 8 %	<i>nähdä</i> 7 %
<i>ajatella</i> 4 %	<i>mainita</i> 6 %
<i>pitää</i> 4 %	<i>käyttää</i> 5 %
<i>sanoa</i> 4 %	<i>selittää</i> 3 %
<i>erottaa</i> 3 %	<i>tutkia</i> 3 %
<i>ilmaista</i> 3 %	<i>kuvitella</i> 2 %
<i>kuvata</i> 3 %	<i>pitää</i> 2 %
<i>jaotella</i> 2 %	<i>rekonstruoida</i> 2 %
[---]	[---]
<b>100 %</b>	<b>100 %</b>

Kuten taulukon 6 jakaumat osoittavat, luokittelua kuvaavat verbit ovat S1-aineistossa selvästi S2-aineistoa yleisempiä passiivisen *voida*-verbin yhteydessä. Edellä esitetyt jakaumat ja huomiot osoittavat kaiken kaikkiaan, että S1-aineistossa *voida*-rakenteita käytetään S2-aineistoa useammin erityisesti passiivissa. Ero näkyy selvimmin jonkinlaista luokittelua kuvaavissa rakenteissa. Luokittelevaa kieltä on aiemminkin pidetty tyyppillisenä nimenomaan tieteelliselle suomelle. Saukkosen mukaan

tieteellis-teoreettiselle tekstille on ominaista järjestyä hierarkkisesti siten, että yläteemalla on alateemoja, ja Saukkosen tieteellis-teoreettista tekstiä kuvaavan esimerkkikatkelman seitsemästä lauseesta kolmessa onkin semanttisena pääverbinä *jakaa*-verbi, joka eksplisiittisesti kielen-tää tällaista hierarkkista suhdetta (Saukkonen 1984: 51, 54).

Tarkastelen seuraavaksi esimerkkien 11 ja 12 avulla hierarkkista jäsentymistä ja sitä, mitä passiivisesta *voida*-verbistä ja merkitykseltään luokittelevasta verbistä koostuva verbiketju ilmaisee ja miten se suhtautuu modaalisuuden eri lajeihin.

- (11) Suomessa verbit **voidaan jakaa** 0-, 1-, 2- tai 3-paikkaisiin verbeihin sen mukaan, moneenko tahoon ne vaikuttavat. --- Suomen verbit **voidaan jakaa** myös hetkellistä ja pidempiaikaista toimintaa kuvaaviin verbeihin. (S1: las2-vtt01vert010)
- (12) 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)

Molemmat esimerkit ilmaisevat jonkin piirteen hierarkkista luokittelua, esimerkissä 11 kuvataan eri tapoja jakaa verbejä alakategorioihin ja esimerkissä 12 kuvataan mordvalaisen eepoksen ryhmittelyä. Molemmat esimerkit ovat tenttivastauskokonaisuuden alussa, ja ne jäsentävät vastausta osiin. Nähdäkseni molemmissa tapauksissa *voida*-verbillä on kaksi toisiinsa kietoutunutta pääasiallista funktiota. Ensinnäkin sillä ilmaistaan sitä, että kuvattava jako on yksi vaihtoehto mutta että muitakin vaihtoehtoja on olemassa. Tämä näkyy eksplisiittisesti esimerkissä 11, kun tekstissä nostetaan esiin vaihtoehtoinen jaottelutapa. Esimerkissä 12 jaottelun valinnaisuutta ilmaisee se, että jaottelun esittelyn jälkeen tekstissä esitellään sellaisia jaosta riippumattomia tekijöitä, jotka ovat yhteisiä kaikille eepoksille. Tämä puolestaan indikoi sitä, että esitetty jako on vain yksi mahdollinen näkemys.

Toinen esimerkeissä näkyvä merkitysfunktio on nähdäkseni se, että verbiketju osoittaa esitetyn tiedon olevan peräisin jostakin lähteestä, niin että se ei välttämättä kuvaa kirjoittajan omaa näkemystä. Kummassakin esimerkissä passiivisen *voida*-verbin ja luokittelua kuvaavan verbin on tenttivastaus-tekstilajissa luontevaa ajatella ilmaisevan likimain merkitystä ‘lukemani kirjan mukaan tai seuraamieni luentojen perusteella X jaetaan...’. Ilmausten sijainti tenttivastauksen alussa tuntuu odotuksenmukaiselta, sillä tekstilajin yhtenä tyyppiirteenä voitaneen pitää sitä, että kirjoittaja osoittaa heti tekstin alussa perustavansa tekstinsä niihin lähteisiin, joihin tekstin odotuksenmukainen vastaanottaja – tentaattori – sen odottaakin perustuvan.

Ilmausten modaalisuuden luonne on monitulkintainen. Tavallaan molemmat ilmaukset ovat episteemisiä – ne kuvaavat kirjoittajan tietoa asiointilan mahdollisesta laidasta (vrt. VISK §1556). Toisaalta se, että ilmaistu mahdollisuus perustunee ennen kaikkea luettuun tenttikirjallisuuteen tai kurssin luentoihin, lähentää tulkintaa kohti dynaamista modaalisuutta, tässä tapauksessa ulkoisista olosuhteista nousevaa mahdollisuutta tai oikeutusta (vrt. VISK § 1554). Tällainen modaalisuuden lajin rinnakkaisuus vastaa Kangasniemen (1992: 392–400) kuvaamaa modaalisuuden lajien yhteensulautumista, eikä ilmausten jakaminen jompaankumpaan luokkaan ole mielekäästä. Molemmat tapaukset kuvaavat kuitenkin kirjoittajan asemaa suhteessa sanottuun: esitetyt jaot ovat kirjoittajien näkökulmasta mahdollisia tapoja jäsentää ilmiötä, mutta eivät välttämättä oikeita, ainoita tai parhaita. Lisäksi ilmaukset etäännyttävät kirjoittajan oman mielipiteen esitetyistä jaosta (vrt. Hyland 1999). Näiden esimerkkien valossa mahdollisuuden ilmauksilla onkin hienovarainen mutta merkityksellinen akateemiselle tekstilajille ominainen funktio, jota käytetään S2-aineistossa S1-aineistoa vähemmän.

## 5. Tulkintoja ja pohdintaa

Olen tässä tutkimuksessa tarkastellut edistyneiden suomenoppijoiden kirjoitetun kielen modaalisuuden piiriin kuuluvia mahdollisuuden ilmauksia. Olen lähestynyt aihetta kahdelta kannalta, koska olen tarkastellut yhtäältä sitä, kuinka paljon eri ilmaisukeinoja käytetään, ja toisaalta sitä, miten ilmaisukeinoista yleisin – *voida*-verbi – tyypillisesti käyttäytyy. Esitän nyt tutkimuskysymyksiini seuraavia vastauksia.

Tulokset osoittavat, että mahdollisuutta ilmaistaan ainakin tutkittujen ilmaisutapojen osalta S2-aineistossa selvästi S1-aineistoa vähemmän – S2-aineistossa ilmauksia on alle 11 jokaista 1000:ta sanaa kohti, kun S1-aineistossa niitä on noin 19 jokaista 1000:ta sanaa kohti. Mahdollisuuden ilmausten sisäinen jakauma puolestaan osoittaa, että eri ilmaisukeinojen osuudet kaikista mahdollisuuden ilmauksista ovat aineistojen välillä hyvin samankaltaiset: noin kolme neljäsosaa tapauksista on *voida*-verbejä, kaikki mahdollisuuden ilmaisukeinot ovat suhteellisesti yleisempiä S1-aineistossa, ja eri ilmaisukeinojen suhteelliset osuudet ovat osa-aineistoissa hyvin samankaltaiset. Tutkimuksen innoittajana oli se osittain samasta aineistosta tehty aiempi huomio, että *voida*-verbistä ja A-infinitiivistä koostuvaa mahdollisuuden ilmausta käytetään S2-teksteissä selvästi S1-tekstejä vähemmän (Ivaska 2014a). Nyt saadut tulokset vahvistavat tämän havainnon ja osoittavat lisäksi, että eroa ei kompensoi mikään muu mahdollisuuden ilmauskeino.

Tutkimuksen toisessa osassa olen tarkastellut yleisimmän mahdollisuuden ilmauksen eli *voida*-verbin tyypillistä käyttöä osa-aineistoissa. Analysoin käyttöä tarkastelemalla kielenpiirteen sisäistä vaihtelua ja sen tyypillisessä kielellisessä lähikontekstissa eli kotekstissa esiintyvää vaihtelua. Saatujen tulosten mukaan käyttö on osa-aineistoissa verraten samankaltaista: *voida*-verbi on yleensä joko aktiivin yksikön kolmannen persoonan preesensissä tai passiivin preesensissä, ja se muodostaa lauseen predikaatin yhdessä jonkin toisen verbin – useimmiten *olla*-verbin – kanssa. *voida*-verbin kanssa tyypillisesti esiintyvien pääverbien ja verbin tyypillisen syntaktisen kotekstin tarkastelu kuitenkin osoittaa sen



eron, että S1-aineistossa *voida*-verbit ovat useammin passiivimuotoisia. Lisäksi S1-aineistossa predikaatin semanttinen pääverbi, erityisesti passiivimuotoisena, kuvaa useammin jonkinlaista luokittelua.

Tulokset ovat monilta osin linjassa aiemmin tehtyjen havaintojen kanssa (vrt. Kemppainen 2001; Niiranen 2008; Seilonen 2013), ja ne tukevat yhdessä aiempien havaintojen kanssa seuraavaa tulkintaa: mahdollisuuden ilmaisemisen keinot ovat edistyneessä oppijansuomessa samankaltaisia kuin ensikielisessäkin suomessa mutta mahdollisuutta ilmaistaan ylipäätään vähemmän. Tyypillisimmät modaalisuuden ilmaisemisen tavat omaksutaan yleensä jo verraten aikaisessa vaiheessa (vrt. Paavola 2008: 48–53), joten rakenteiden tietoinen välttäminen ei itsessään selitä frekvenssieroja. Katsonkin aineistojen välisen eron juontavan juurensa nimenomaan tekstilajille tyypillisiin hienovaraisiin funktioihin, joita edistyneetkään oppijat eivät välttämättä hallitse ensikielisten kaltaisesti.

Akateemiseen kieleen kuuluu inherentisti erilaisten mahdollisuuksien esittäminen. Usein käsiteltävästä asiasta voi olla toisistaan poikkeavia näkökantoja, minkä voidaan ajatella lisäävän mahdollisuuden ilmaisemista kyseisessä tekstilajissa. Aineistoni koostuu tenttivastauksista, jotka ovat jo perusluonteeltaan sellaisia, että niissä esitetään jonkun muun kuin kirjoittajan omia näkemyksiä. Usein modaalisisista tapauksista välittyikin ajatus siitä, että esitetyn näkemyksen lisäksi on myös muita vaihtoehtoisia selityksiä, ja kirjoittaja etäännyttää oman mielipiteensä näistä näkökulmista. Osa-aineistojen välinen käyttöero lienee tyypillinen esimerkki kontekstuaalisesta epäkonventionaalisuudesta (Jantunen 2008: 81–84). Kuten Jantunen huomauttaa “kolligaatioihin ja kotekstin semanttisiin piirteisiin ei ole kiinnitetty huomiota tähänastisessa tutkimuksessa. --- [S]yntagmaattiset kieliopilliset ja semanttiset valinnat ohjaavat lekseemien käyttöä yhtä merkittävästi kuin kollokationaalisetkin valinnat.” (mt. 83). Tuntuukin luontevalta ajatukselta, että mahdollisuuden ilmaisemisen tekstilajikohtaiset käytänteet akateemisessa kirjoituksessa edustavat tällaista kontekstisidonnaista syntagmaattista valintaa (vrt. Paavola 2008: 50–51). Näin ollen havaitun eron taustalla

on nimenomaan tutkitulle tekstilajille tyypillinen kielellinen keino ja se, että suomea toisena kielenään kirjoittavat hyödyntävät tätä keinoa ensikielisiä kirjoittajia harvemmin.

Monet mahdollisuuden ilmaukset kuvaavat akateemisessa kielessä kirjoittajan omaa positiota suhteessa aiemmin esitettyyn, käytännössä jonkun muun esittämään (vrt. Hyland 1999). Vastaavasti myös suomen passiivilla kuvataan tyypillisesti toimintaa yleisellä tasolla jättämällä tekijä taka-alalle niin, että tekijä kuitenkin on yleensä inhimillinen (VISK § 1315). Siksi onkin mielenkiintoista, että ero ei näytä kielikohdaiselta opittavan kielen tai oppijan ensikielen kannalta. Tämän tutkimuksen S2-aineisto on luonteeltaan moniensikielinen, ja ilmiö toistuu hyvin samankaltaisena halki aineiston. Vastaava tendenssi on havaittu akateemisen oppijansuomen lisäksi myös akateemisen oppijanenglannin samankaltaisessa aineistossa (vrt. Cotton 2010). Ilmiö ei siis selity kieltenvälisillä vaikutussuhteilla eikä kohdekielen ominaisuuksilla, vaan syynä on oltava toisella kielellä toimiminen tietyssä tekstilajissa. Tämä vastaa myös Maurasen (1996) aiempaa huomiota siitä, että tekstilajikohtaisten rakennepiirteiden hallinta (engl. *discourse construction ability*) on ennen kaikkea kielitaitokysymys.

Cottonin (2010) arvelun mukaan kielenoppijat käyttävät suhteessa vähemmän modaali-ilmauksia todennäköisesti siksi, että he eivät vielä kunnolla hallitse kriittiseen kielenkäyttöön soveltuvaa kielirekisteriä. Ajatus tuntuu perustellulta myös kompleksisuusteorian ja käyttöperusteiden mallien näkökulmasta: Vaikka sekä S1- että S2-informantit ovat olleet tekemisissä akateemisen kielen kanssa, S2-kielenkäyttäjille akateeminen suomi on kuitenkin vieraampi kielellinen konteksti kuin ensikielisille. Valtaosa tämän tutkimuksen S2-informanteista on opiskellut alemman korkeakoulututkintonsa jossain muualla kuin Suomessa, joten suomalaiset ja suomenkieliset akateemisen kielen konventiot ovat heille vieraampia. Lisäksi monille koko tenttivastaus kirjoitettuna tekstilajina saattaa olla melko vieras – useissa maissa pääosa tenteistä suoritetaan suullisina. Käytettävät konstruktiot ovat kielenoppijoille sinänsä tuttuja, mutta niitä käytetään osin eri tavoin uudessa, oppijoille osin vieraassa

käyttöympäristössä. Tämä varmaan osaltaan osoittaa, miten frekvensitekijät vaikuttavat dynaamisesti: modaali-ilmausten käyttäminen kirjoittajaposition ilmaisukeinona tässä tutkimuksessa tarkastellussa tekstilajissa ei edistyneiden oppijoiden kielessä ole vielä vakiintunut ainakaan samassa laajuudessa kuin ensikielisillä kielenkäyttäjillä (vrt. N. Ellis & Larsen-Freeman 2006). Dynaamisena tekijänä saattaa siis tässä tapauksessa olla ainakin osittain uusi tekstilaji – akateeminen suomi ja sen alalaji tenttivastaus.

Tekstilajikohtaisten lingvististen erojen ja tyyppipiirteiden kohdekielimäinen hallinta ja käyttö lienee yksi hyvin edistyneen kielenoppijan osoittimista. Tässä artikkelissa tutkitut mahdollisuuden ilmaukset kuvaavat akateemisissa teksteissä usein kirjoittajan omaa asemaa suhteessa esitettyyn sisältöön, mikä on semanttisesti hienovarainen merkitysvivahde. Tällaiset tekstilajikohtaiset toisen kielen käyttäjiä ensikielisistä erottavat seikat tarjoavat hedelmällisiä tilaisuuksia tarkastella kielijärjestelmän dynaamista ja avointa luonnetta. Ilmiötä olisikin mielenkiintoista tarkastella myös pitkittäismuutoksen osalta – tässä artikkelissa tehtyä tulkintaa tukisi, jos voitaisiin osoittaa mahdollisuuden ilmausten frekvenssin kasvavan sitä mukaa, kuin kirjoittaja jatkaa toimintaansa akateemisessa kontekstissa. Olen tässä tutkimuksessa verrannut tuloksia lähinnä akateemisesta englannista tehtyihin havaintoihin. Ilmiötä olisi kuitenkin mielenkiintoista tarkastella myös monikielisestä asetelmasta niin, että sekä lähtökieliä ja -kulttuureita että kohdekieliä ja -kulttuureita olisi useampia. Ilmiön toistuminen samankaltaisena vahvistaisi sekin puolestaan tässä artikkelissa esitettyjä näkemyksiä. Toivonkin, että tämä tutkimus innostaa myös uusien tutkimusasetelmien rakentamiseen ja akateemisen suomen tyyppipiirteiden ja niiden oppimisen tutkimukseen nykyistä laajemmin.

**Lähteet**

- Bates, E., B. MacWhinney 1987. Competition, variation and language learning. – Brian MacWhinney (Ed.). *Mechanisms of Language Acquisition*. Hillsdale (NJ): Lawrence Erlbaum Associates, 157–193.
- Biber, D. 2006. *University Language. A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/scl.23>
- Bybee, J., S. Thompson 1997. Three frequency effects in syntax. – *Berkeley Linguistics Society* 23, 65–85.
- Bley-Vroman, R. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. – *Language Learning* 33 (1), 1–17. <http://dx.doi.org/10.1111/j.1467-1770.1983.tb00983.x>
- Campbell, P. N. 1975. The *personae* of scientific discourse. – *Quarterly Journal of Speech* 61 (4), 391–405. <http://dx.doi.org/10.1080/00335637509383302>
- CEFR = Common European Framework for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press, 2006.
- Cotton, F. 2010. Critical thinking and evaluative language use in academic writing: A comparative cross-cultural study. – G. Blue (Ed.). *Developing Academic Literacy*. Oxford: Peter Lang, 73–85.
- Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780198299554.001.0001>
- Cowie, A. P. 1981. The treatment of collocations and idioms in learners' dictionaries. – *Applied Linguistics* 2 (3), 223–235. <http://dx.doi.org/10.1093/applin/2.3.223>
- Ellis, N., D. Larsen-Freeman 2006. Language emergence: Implications for applied linguistics. – *Applied Linguistics* 27 (4), 558–589. <http://dx.doi.org/10.1093/applin/aml028>
- Ellis, R. 1985. *Understanding Second Language Acquisition*. Oxford: Oxford University Press.
- Fillmore, C. J. 1985. Syntactic intrusions and the notion of grammatical construction. – *Berkeley Linguistics Society* 11, 73–86.
- Flint, A. 1980. *Semantic Structure in the Finnish Lexicon: Verbs of Possibility and Sufficiency*. SKS:n toimituksia 360. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Francis, G. 1993. A corpus-driven approach to grammar – principles, methods and examples. – M. Baker, G. Francis, E. Tognini-Bonelli (Eds.). *Text and*

- Technology. In Honour of John Sinclair. Amsterdam: John Benjamins, 137–156.
- Goldberg, A. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, A. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, A., D. M. Casenhiser, N. Sethuraman 2004. Learning argument structure generalizations. – *Cognitive Linguistics* 15 (3), 289–316. <http://dx.doi.org/10.1515/cogl.2004.011>
- Granger, S. 1996. From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. – K. Aijmer, B. Altenberg, M. Johansson (Eds.). *Languages in Contrast*. Lund: Lund University Press, 37–51.
- Granger, S. 1998. Prefabricated patterns in advanced EFL writing: Collocations and formulae. – A. P. Cowie (Ed.). *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press, 145–160.
- Granger, S. 2013. Contrastive interlanguage analysis: A reappraisal. – Esitelmä konferenssissa Learner Corpus Research 2013. Bergen 27.–29.9.2013.
- Gries, S. Th. 2008a. Phraseology and linguistic theory. – S. Granger, F. Meunier (Eds.). *Phraseology: An Interdisciplinary Perspective*. Amsterdam: John Benjamins, 3–25.
- Gries, S. Th. 2008b. Dispersions and adjusted frequencies in corpora. – *International Journal of Corpus Linguistics* 13 (4), 403–437. <http://dx.doi.org/10.1075/ijcl.13.4.02gri>
- Hollander, M., D. A. Wolfe 1973. *Nonparametric Statistical Methods*. New York: John Wiley & Sons.
- Hopper, P. 1987. Emergent grammar. – *Berkeley Linguistics Society* 13, 139–157.
- Hyland, K. 1999. Disciplinary discourses: Writer stance in research articles. – C. N. Candlin, K. Hyland (Eds.). *Writing: Texts, Processes and Practices*. London: Longman.
- Inaba, N. 2007. Mikael Agricolan teokset tietokannan muodossa. – K. Häkkinen, T. Vaittinen (Toim.). *Agricolan aika*. Helsinki: BTJ, 147–161.
- Itkonen, E. 2005. *Analogy as Structure and Process. Human Cognitive Processing* 14. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/hcp.14>
- Ivaska, I. 2012. Keystructure analysis of formally defined structures of learner Finnish. – Esitelmä konferenssissa Learner Language, Learner Corpora. Oulu 5.–6.10.2012.
- Ivaska, I. 2014a. Edistyneen oppijansuomen avainrakenteita. Korpusnäkökulma kahden kielimuodon tyypillisiin rakenteellisiin eroihin. – *Virittäjä* 108 (2), 161–193.

- Ivaska, I. 2014b (tulossa). The Corpus of Advanced Learner Finnish (LAS2): Database and Toolkit to Study Academic Learner Finnish. – Apples 8 (3).
- Jantunen, J. H. 2004. Synonymia ja käännössuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin. Joensuun yliopiston humanistisia julkaisuja 35. Joensuu: Joensuun yliopistopaino.
- Jantunen, J. H. 2008. Haasteita oppijankielen korpusanalyysille: oppijankielen universaalit. – Pille Eslon (Toim.). Öppijakeele analüüs: Võimalused, probleemid, vajadused. Tallinna Ülikooli eesti filoloogia osakonna toimetised 10. Tallinn: Tallinna Ülikool, 67–92.
- Jantunen, J. H. 2011. Kansainvälinen oppijansuomen korpus (ICLFI): typologia, taustamuuttujat ja annotointi. – Lähivõrdlusi. Lähivertailuja 21, 86–105. <http://dx.doi.org/10.5128/LV21.04>
- Jantunen, J. H. 2012. Akateemiset ja populaaritekstit korpusvertailussa: sanalistat, avainsanat ja fraseologiset yksiköt. – V. Heikkinen, E. Voutilainen, P. Lauerma, U. Tiililä, M. Lounela (Toim.). Genreanalyysi – tekstilajituskimuksen käytäntöä. Kotimaisten kielten keskuksen verkkojulkaisuja 29. Helsinki: Kotimaisten kielten keskus, 186–213. <http://kaino.kotus.fi/www/verkkojulkaisut/julk29/> (2.9.2014).
- Kangasniemi, H. 1992. Modal Expressions in Finnish. *Studia Fennica Linguistica* 2. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kemppainen, M. 2001. Modaalisuus suomea toisena kielenä oppivien kirjoitelmissa. Pro gradu -tutkielma. Tampereen yliopiston suomen kielen laitos.
- Laalo, K. 2011. Lapsen varhaiskielioppi ja miniparadigmat. SKS:n toimituksia 1309. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Larsen-Freeman, D. 1997. Chaos/complexity and second language acquisition. – *Applied Linguistics* 18 (2), 141–165. <http://dx.doi.org/10.1093/applin/18.2.141>
- Larsen-Freeman D., L. Cameron 2008. *Complex Systems and Applied Linguistics*. Oxford: Oxford University Press.
- Levenston, E. A. 1971. Over-indulgence and under-representation: Aspects of mother tongue interference. – G. Nickel (Eds.). *Papers in Contrastive Analysis*. Cambridge: Cambridge University Press, 115–121.
- Lieven, E., J. Pine, G. Baldwin 1997. Lexically-based learning and early grammatical development. – *Journal of Child Language* 24 (1), 187–219. <http://dx.doi.org/10.1017/S0305000996002930>
- Mauranen, A. 1993. *Cultural Differences in Academic Rhetoric. A Textlinguistic Study*. Frankfurt am Main: Peter Lang.

- Mauranen, A. 1996. Discourse competence – evidence from thematic development in native and non-native texts. – E. Ventola, A. Mauranen (Eds.). *Academic Writing: Intercultural and Textual Issues*. Amsterdam: John Benjamins, 193–230.
- Meunier, F., Granger, S. (Eds.) 2008. *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: Benjamins. <http://dx.doi.org/10.1075/z.138>
- Nesselhauf, N. 2004. *Collocations in a Learner Corpus*. SCL 14. Philadelphia: John Benjamins.
- Niiranen, L. 2008. *Effects of Learning Contexts on Knowledge of Verbs. Lexical and Inflectional Knowledge of Verbs among Pupils Learning Finnish in Northern Norway*. PhD dissertation. University of Tromsø.
- Osborne, J. 2013. Comparisons are odorous: Native-speaker data in learner corpus research. – Esitelmä konferenssissa *Learner Corpus Research 2013*. Bergen 27.–29.9.2013.
- Paavola, V. 2008. *Haluaisitko mennä muunkansa kalastaman? Verbiketjujen kehkeytyminen suomi toisena kielenä -oppijoiden kielessä*. Pro gradu -tutkielma. Jyväskylän yliopiston kielten laitos, suomen kieli.
- R 2013 = R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Saukkonen, P. 1984. *Mistä tyyli syntyy?* Juva: WSOY.
- Seilonen, M. 2013. *Epäsuora henkilöön viittaaminen oppijansuomessa*. Jyväskylä Studies in Humanities 197. Jyväskylä: Jyväskylä University Printing House.
- Seppälä, T. 2012. *Oppijansuomen kolligaatit ketjuuntuuvissa verbirakenteissa*. Pro gradu -tutkielma. Oulun yliopiston suomen kielen oppiaine.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Skousen, R. 1989. *Analogical Modeling of Language*. Dordrecht: Kluwer Academic Publishers.
- Stefanowitsch, A., S. Th. Gries 2003. *Collostructions: Investigating the interaction of words and constructions*. – *International Journal of Corpus Linguistics* 8 (2), 209–243. <http://dx.doi.org/10.1075/ijcl.8.2.03ste>
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. *Studies in Corpus Linguistics* 6. Amsterdam: Benjamins. <http://dx.doi.org/10.1075/scl.6>
- Tomasello, M. 1992. *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511527678>

- Tomasello, M. 2003. *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge (MA): Harvard University Press.
- VISK = A. Hakulinen, M. Vilkuna, R. Korhonen, V. Koivisto, T. R. Heinonen, I. Alho 2004. *Iso suomen kielioppi*. SKS:n toimituksia 950. Helsinki: Suomalaisen Kirjallisuuden Seura. <http://scripta.kotus.fi/visk> (2.9.2014). URN:ISBN:978-952-5446-35-7.



## Expressions of possibility as a distinguishing feature between L1-Finnish and advanced L2-Finnish

ILMARI IVASKA

University of Turku

The purpose of this article is to investigate expressions of possibility in the Finnish language in texts written by advanced non-native speakers of Finnish (F2) and compare them with those written by native speakers of Finnish (F1). The aim of the study is to explore how possibility is typically expressed and to find out if or how the compared language varieties differ from each other. Particular emphasis is given to the use of the verb *voida* ‘to may, to be able to,’ which can be used to express many different kinds of modal meanings. The data is a part of the Corpus of Advanced Learner Finnish (LAS2), and the F2-subjects are typically at the B2 or C1 level in the Common European Framework for Languages (CEFR). All the texts included in this study are exam essays and, thus, the results describe typical language use in a specific genre of academic writing.

F1- and F2-users express possibility similarly in many ways, and the distributions between different expressions are very similar between the data sets. However, F1-users tend to express possibility twice as much as the F2-users do. The verb *voida* is the most frequent of all the expressions of possibility as it accounts for three fourths of all the occurrences. Despite the similar patterns of use between the data sets, the occurrences of the verb *voida* are more often in passive voice in the F1 data set than they are in the F2 data set. Furthermore, in the F1 data set the verb *voida* is more often used together with verbs expressing classification or hierarchical structuring of knowledge, which is in turn considered to be characteristic of academic style.

Modal verbs are often used to express stance, i.e. to self-project the writer into their own texts and to communicate their positioning and relationship to the subject matter. Stance is particularly typical for academic language. The difference in the usage is not due to the novelty of the semantic concept or the construction as such, as it is typically acquired earlier on in the learning process. The differences found between the data sets may be due to contextual unconventionality.

In other words, non-native speakers of Finnish do not express a semantic detail that is particularly typical to a specific genre as frequently as native speakers do.

**Keywords:** modality; Finnish as a second language; verb chain; academic language; corpus; contrastive interlanguage analysis

**Ilmari Ivaska**

20014 Turun yliopisto, Finland

[itivas@utu.fi](mailto:itivas@utu.fi)