

Morfologinen priming ja fraseologia vieraan kielen oppimisessa: korpustutkimus oppijansuomesta

JARMO HARRI JANTUNEN

Oulun yliopisto

SISKO BRUNNI

Oulun yliopisto

Tiivistelmä. Artikkelimme käsittelee morfologista primingia eli taivutusmuotojen käyttöä ja muotoihin liittyviä kontekstuaalisia piirteitä. Oletuksemme on, että morfofonologisesti rikkaissa kielessä morfologinen priming on olennainen osa kielen fraseologia-luonnetta, minkä vuoksi ehdotamme hypoteesia morfologisesta primingista osaksi Michael Hoeyn postulaatioita leksikaalisesta primingista (engl. *lexical priming*). Pilottiluonteinen tarkastelumme keskittyy lekseemeihin IHMINEN ja PITÄÄ¹. Tutkimusaineistona käytämme Oulun Kansainvälistä oppijansuomen korpusta (ICLFI). Analyysi perustuu korpusvetoiseen lähestymistapaan (avainsana-analyysiin), ja itse tutkimus on tehty kolmivaiheisella vertailumenetelmällä (engl. *Three-Phase Comparative Analysis*). Tulokset osoittavat, että morfologinen priming ja fraseologia ovat syy-seuraussuhteessa toisiinsa: kielenoppijoiden tuottama epätyypillinen fraseologisuus selittää osan natiivikiielestä poikkeavasta morfologisesta primingista. Lisäksi käy ilmi, että morfologinen priming on sidoksissa myös polyseemisen sanan eri merkityksiin. Odotuksenvastaisesti lähisukukieltä puhuvat eivät kuitenkaan tuota virheittämpiä taivutusmuotoja kuin kielenoppijat yleensä. Morfologisen primingin tarkastelu paljas-

¹ Tarvittaessa kaikki taivutusmuodot sisältävä abstraktio eli lemma on merkitty kapiteelein, sananmuodot puolestaan kurssiivein.

tuu lopulta monisyiseksi, sillä tuotosten taustalla voivat vaikuttaa kieltenvälisen vaikutuksen lisäksi monet muutkin syyt.

Avainsanat: morfologinen priming; leksikaalinen priming; fraseologia; kieltenvälinen vaikutus; oppijansuomi; korpustutkimus

1. Johdanto

Suomen kielen oppimisen ja oppijansuomen tutkimus on viime vuosina haarautunut moniaalle. Yksi 2000-luvun ensi vuosikymmenen tulokas on oppijansuomen korpustutkimus; sille on viitoittanut tietä 20-vuotinen perinne oppijanenglannin korpuspohjaisessa tutkimuksessa. Englantia opiskelevien oppijoiden tuottaman kielen korpustutkimus pohjaa pitkälti firthiläis-sinclairiläiseen traditioon, jossa näkökulmana on erityisesti kielenkäyttö luonnollisissa tilanteissa. Tutkimukseen on lisäksi yhdistetty usein virheanalyysi ja oppijankielen (välikielen) vertailu natiivikieleen; Granger (1996) on käyttänyt näiden menetelmien kombinaatiosta nimeä *Contrastive Interlanguage Analysis*. Oppijansuomen tutkimukseen menetelmä ei ole toistaiseksi rantautunut sellaisenaan, ja tutkimuskysymyksetkin ovat muotoutuneet osittain erilaisiksi – esimerkiksi virheanalyysi ei ole noussut oppijansuomen korpustutkimuksessa samalla tavalla keskiöön.

Käsillä olevalla tutkimuksella on yhteyksiä Grangerin johtaman tutkimushankkeen keskeisiin tavoitteisiin eli oppijankielen fraseologian tutkimukseen. Useissa hankkeen tutkimuksissa (Granger 1998; Nesselhauf 2003; 2005; Meunier & Granger 2008) on havaittu, että kielenoppijoiden tuotokset poikkeavat natiivituotoksista selvästi juuri fraseologisten piirteiden osalta. Fraseologisilla piirteillä tarkoitetaan tutkimuksissa usein kollokaatioita, vaikka fraseologisuuteen lasketaan kuuluvaksi – Sinclairin (1991) idiomiperiaatteen mukaisesti – myös tekstikontekstin kieliopilliset ja semanttiset tasot (ks. Jantunen 2009b). Koska tutkimus on keskittynyt nimenomaan englannin ja joidenkin muiden indoeurooppalaisten kielen tutkimukseen, eivät morfologiset piirteet ole juurikaan saaneet huomiota (natiivi- ja) oppijankielen

corpustutkimuksessa. Taivutukseltaan monipuolisen kielen oppimista koskevassa tutkimuksessa morfofonologisten piirteiden tarkastelu osana kielen fraseologisuutta on kuitenkin olennaista: fraseologisen yksikön ydin ja kontekstuaaliset elementit suosivat yleensä tiettyjä taivutusmuotoja. Tässä artikkelissa pyritään kuvaamaan fraseologisten yksiköiden ja morfofonologisen vaihtelun yhteenkuuluvuutta kahden tapaustutkimuksen avulla: korpusanalyysin perusteella tutkittaviksi tapauksiksi valikoituivat nomini IHMINEN ja verbi PITÄÄ. Niiden avulla on tarkoitus testata hypoteesia ns. morfofonologisen primingin merkityksestä fraseologisudessa ja kielenoppimisessa, joskaan mitään yleistä kuvausta morfofonologian ja fraseologian keskinäisestä suhteesta oppijankielessä (tai natiivikielissä) ei tapaustutkimusten perusteella voikaan antaa.

2. Leksikaalinen priming ja fraseologia

Psykolingvistiikan – ja laajemminkin psykologian, muistin ja oppimisen tutkimuksen – käsite priming tarkoittaa ärsykkeeseen tai ennakkotietoon perustuvaa ennakkointia, virittäytymistä ja reagointia. Kun henkilö havaitsee jonkin ärsykkeen, esimerkiksi kuvan tai sanan, havaittu ärsyke aktivoi siihen liittyviä aiemmin koettuja tapahtumia, kokemuksia, käsitteitä, sanoja tai ilmauksia. Näin ollen aiemmin havaittu ja opittu vaikuttaa myöhempään oppimiseen ja esimerkiksi kielellisiin tuotoksiin. Aiemmat kokemukset säilyvät implisiittisessä muistissa, josta ärsyke aktivoi ja joka helpottaa muun muassa uuden oppimista ja mieleen palauttamista. Mitä useammin samanlainen ärsyke on aktivoinut tiettyntyyppisen reaktion, sitä helpommin odotuksenmukainen priming toteutuu; niinpä esimerkiksi länsimaisen ihmisen ei syödessään tarvitse erikseen miettiä, miten haarukkaa käytetään, vaan tieto on hänen implisiittisessä muistissaan ja aivot ovat valmistautuneet tuottamaan tietyn prosessoinnin. Vastaavasti syömäpuikkoihin tottunut ei joudu erityisesti pohtimaan niiden käyttämistä.

Ärsykkeen tuottamasta aktivoitumisesta käytetään käsitettä priming-efekti. Psykolingvistisessä tutkimuksessa on tarkasteltu semant-

tista virittäytymistä erityisesti sanaparitesteillä (ks. mm. Neely 1991; McNamara 2005): kun sanaparin sanat kuuluvat samaan merkityskenttään, ensimmäinen sana aktivoi onnistuneesti ja nopeasti seuraavan sanan, mutta aivan eri merkityskenttään kuuluva sana puolestaan häiritsee ja hidastuttaa jälkimmäisen sanan tuottamista tai tunnistamista. Niinpä esimerkiksi sanojen tunnistamistestissä (epäsanat vs. todelliset sanat) todellisista sanoista tunnistetaan nopeammin ne, joita edeltävä ärsyke on samaan merkitysryhmään kuuluva (engl. *nurse–doctor*), kuin ne, joita edeltää toiseen ryhmään kuuluva sana (engl. *bread–door*) (McNamara 2005: 3). Tunnistamisen helppous riippuu merkitysten läheisyydestä ja siitä, kuinka taajaan sanat esiintyvät yhdessä.

Michael Hoey on puolestaan tehnyt tunnetuksi käsitteen leksi-kaalinen priming (engl. *lexical priming*). Tämä käsite on kuitenkin pidettävä erillään psykolingvistiikan priming-käsitteestä, joka liittyy edellä kuvatulla tavalla ärsykkeen aikaansaamaan reaktioon koetilanteessa. Näin ollen tutkimaamme ilmiötä ei voi pitää psykolingvistiseen priming-tutkimukseen kytkeytyvänä, vaan se kuvaa kielen fraseologista luonnetta. Siinäkin on kyse eräänlaisesta virittäytymisestä sanojen tyyppilliseen käyttöön, mutta priming on kuitenkin määritelty hieman toisin tarkoittamaan kielenomaksumisen aikana kielitietoon muodostunutta skeemaa sanojen tyyppillisestä käytöstä:

--- when we use the word, we will characteristically use it with one of its typical collocations, in its usual grammatical function, in the same semantic context, in the domain we have come to associate it with, as part of the same genre, in a familiar social context, with a similar pragmatics and to similar textual ends. (Hoey 2007: 8.)

Hoeyn näkemys sanojen tyyppillisten käyttöyhteyksien hallinnasta on siis varsin holistinen, ja se kattaa Sinclairin (1996) määritelmän merkityksiköstä ytimineen ja sen kontekstuaalisine valintasuhteineen. Priming liittyy siis laajasti fraseologiaan: ilmauksilla on odotuksenmukaiset käyttöyhteytensä alkaen tyyppillisistä kerasanoista eli kollokaateista aina tyyllisiin ja tekstuaalisiin preferensseihin asti. Käyttöyhteyksien omak-

sumisessa kertautumisella on erittäin suuri merkitys. Samalla tavalla kuin sanojenkin omaksuminen perustuu toistumiseen, myös kontekstuaalisten suhteiden hallintaan tarvitaan prosessi, jossa tietyt rakenteet toistuvat taajaan. Niinpä natiivi kielenkäyttäjä “tietää”, että *voi härskiintyy* tai *eltaantuu*, *leipä* ja *hillo homehtuvat*, mutta *maito* puolestaan *happanee*. Härskiintyä voivat *voin* lisäksi myös *öljyt*, *margariinit*, *rasvat* ja esimerkiksi runsaasti rasvaa sisältävät kalat, kuten *lohet*. Niinpä rasvahappojen hapettumista tarkoittavalla *härskiintymisellä* on semanttinen preferenssi (ks. esim. Jantunen 2009b) rasvoja tai niitä runsaasti sisältäviä ruoka-aineita merkitseviin sanoihin. Tämä näkyy siinä, että kielenkäyttäjät käyttävät verbiä tyypillisesti tällaisten sanojen kanssa. Natiivin kielenkäytössä priming näyttäytyy siis juuri toistuvina käyttöyhteyksinä. Sanojen fraseologisuuteen kuuluu runsaasti erilaista konkreettista ja abstraktia tietoa sen käytöstä; tässä yhteydessä fraseologisuudella tarkoitetaan siis huomattavasti laajempaa sanastollis-kieliopillis-semanttista ja tekstuaalista ilmiötä kuin yleensä puhuttaessa esimerkiksi idiomeista, metaforisista ilmauksista tai sanonnoista.

3. Kielenoppiminen ja leksikaalinen priming

Kun S2-kielenkäyttäjä tuottaa lauseen *Ulkona ei ole erittäin kaunis ilma*, on hän muodostanut kieliopillisesti kelvon rakenteen – ja huomattakoon, että myös ymmärtämisen näkökulmasta onnistuneen ilmauksen. Fraseologisuuden näkökulmasta se on kuitenkin epätyypillinen: *erittäin* ei kollokoi kieltoverbin kanssa, eikä sillä ole negatiivista semanttista prosodiaa. Tällä hetkellä tutkimuskirjallisuudessa (ks. esim. Wray 1999; 2000; Nesselhauf 2005; Tsui 2004; 2005; Ellis 2008; Meunier & Granger 2008) ollaan yksimielisiä siitä, että kohdekielen fraseologisuus on yksi vaikeimmin omaksuttavista ja opetettavista asioista, mutta natiivin kaltaisen kielenkäytön tavoittamisen kannalta sen hallinta on kuitenkin tärkeää. Fraseologisuuden ymmärtäminen ja kyky muodostaa kohdekielen mukaisia ilmauksia ovat nimittäin yhteydessä puhutun ja kirjoitetun kielen sujuvaan käyttöön. Lisäksi syötöksen ymmärtäminen nopeutuu

ja helpottuu, kun viestin vastaanottaja pystyy hahmottamaan syötöksestä tuttuja, valmiita kokonaisuuksia sen sijaan, että jokaisen yksittäisen ilmauksen merkitys ja suhde toisiin ilmauksiin pitäisi analysoida itsenäisinä ja erikseen. Reseptiivisen ja produktiivisen sujuvuuden kehittymisen lisäksi fraseologinen tieto auttaa myös rekisterien ja tekstityyppien hallinnassa, sillä niiden konventioihin kuuluu oleellisena osana omanlaisensa fraseologisuus.

Leksikaalinen priming säilyy kielenoppijan ongelmana pitkään, sillä vielä hyvin edistyneetkin kielenoppijat tuottavat kohdekielen fraseologisesta normistosta poikkeavia muotoja (Granger 1998; Hasselgren, 1994; Nesselhauf 2005; 2009). Luultavaa on, että taajaan esiintyvät ja morfosyntaktisesti selvästi vakiintuneet fraseologiset rakenteet, kuten keskustelurutiinit, omaksutaan helpommin kuin (ainakin teoreettisesti) vapaat yhdistelmät. Tähän asti fraseologisuuden problematiikassa on ollut tutkimuksen kohteena nimenomaan kollokaatioiden hallinta ja muiden piirteiden analyysi on jäänyt mitättömäksi. Kielenoppijoiden tuotoksista on esimerkiksi havaittu, että oppijat keskittyvät käyttämään tiettyjä frekventtejä kollokaatioita, jolloin reseptiivinen ja produktiivinen käyttö helpottuvat näiltä osin, mutta kohdekielessä ilmenevä kollokaatiokirjo jää oppijoilta tavoittamatta. Toisaalta oppijat tuottavat runsaasti myös kohdekielestä poikkeavia kollokaatioita. Tähän ovat syynä yhtä hyvin kollokaationormiston hallinnan puute kuin myös kielenoppijan luontainen tarve pyrkiä luovaan ja kokeilevaan kielenkäyttöön. (Granger 1998; Kaszubski 2000; Nesselhauf 2005.) L2-oppijoiden kollokaatioita on tarkasteltu korpusaineistojen avulla jo myös suomen kielen oppijoiden tuotoksista (ks. Kallioranta 2009), mutta tutkimustieto koostuu edelleen lähinnä englannin oppimisesta tehdyistä havainnoista. Yhtä kaikki fraseologisuuden tutkimus on raapaissut toistaiseksi vain pintaa, sillä monet sen ilmiöt – kuten tässä artikkelissa analysoitava morfologinen priming – ovat jääneet vielä huomiotta. Lisäksi huolimatta siitä, että kollokaatioiden oppimista on analysoitu runsaasti, ei niiden (saati muiden fraseologisten piirteiden) opettamisen metodisessa kehittämisessä ole päästy vielä kovinkaan pitkälle – ei edes englantia opettaessa.

Kollokaatioiden oppimisen tutkimuksessa on keskitytty pitkälti kieltenoppijoiden tuotoksen ja natiivituotoksen vertailuun sen sijaan, että opetuksen tueksi olisi selvitetty natiivien tuotoksissa ilmeneviä säännönmukaisuuksia. (Nesselhauf 2005: 3, 6.)

4. Morfologinen priming

Kun tarkastellaan morfofonologisesti rikkaiden kielten, kuten suomen, fraseologiaa, kohtaa analyysi haasteita: pitäydytäänkö fraseologisen yksikön analyysissa lemmatasolla eli abstraktissa perusmuodossa (kuten IHMINEN vs. *ihminen, ihmisen, ihmistä...*) vai otetaanko tarkasteluun mukaan myös taivutusmuodot ja niissä ilmenevä variaatio (esim. *ihmisten* vs. *ihmisien*). Aiempi, erityisesti englantia koskeva fraseologiatutkimus ei ole juuri kiinnittänyt luonnollisista syistä taivutukseen kovinkaan paljon huomiota. Suomen kielessä taipuminen on kuitenkin merkittävä ominaisuus sekä fraseologisen yksikön ytimen (engl. *core*) että sen kotekstuaalisten kerailmausten näkökulmasta: molemmathan taipuvat morfosyntaktisten vaatimusten ja rajoitusten mukaisesti, eli niillä on tietty odotuksenmukainen priming johonkin taivutusmuotoon. Niinpä natiivin suomen kielen puhujan priming ei toteudu pelkästään tiettyinä kotekstuaalisina leksikaalisina ja semanttisina valintoina vaan myös itse ytimen ja sen kerailmausten taivutusmuotoina. Tätä ilmiötä kutsumme seuraavassa esityksessä fraseologisen yksikön **morfologiseksi primingiksi**. Aiemmin ilmiöstä on käytetty myös nimitystä morfologinen preferenssi (Jantunen 2004a), joka jakautuu paradigmaattiseen ja kontekstuaaliseen taivutusmuotojen preferenssiin. Käsillä olevassa artikkelissa käytämme kuitenkin käsitettä morfologien priming, koska haluamme liittää ilmiön laajempaan leksikaalisen primingin käsitteeseen. Ilmiöön on viitannut aiemmin myös Hoey (2005: 159) primingin ja kieliopin suhdetta käsitellessään, mutta teoria leksikaalisesta primingista ei kuitenkaan ota fraseologisen yksikön morfologiaa piirteitä huomioon. Oletuksemme kuitenkin on, että morfologisen primingin käsitteleminen on oleellista fraseologisia yksiköitä

tutkittaessa; niinpä esitämmekin Hoeyn (2005: 13) hypoteeseja mukailleen seuraavaa:

Jokaisella sanalla on oma morfologinen primingsinsa, eli se esiintyy tyypillisesti ja frekventisti yhdessä tai useammassa morfofonologisessa muodossa, ja myös sen kontekstissa esiintyvistä ilmauksista on löydettävissä omat odotuksenmukaiset taivutusmuotonsa.

Oletus tiettyjen taivutusmuotojen tyypillistymisestä ei ole tietystikään uusi: on yhtäältä esitetty, että sanalla voi olla käytössä vain tiettyjä taajaan käytettyjä taivutusmuotoja, ja toisaalta, että tietyillä taivutusmuodoilla voi olla oma kontekstuaalinen preferenssinsä. Ensimmäisestä tapauksesta käy esimerkiksi synonyymipari TÄRKEÄ ja KESKEINEN, joista TÄRKEÄ esiintyy eniten partitiivimuotoisena ja melko taajaan komparoituna mutta sen synonyymi puolestaan nominatiivissa ja yleensä positiivimuotoisena (Jantunen 2001). Jälkimmäisestä antaa esimerkin Tognini-Bonelli (1996: 74, 77–80): italian *tietää*-verbin vastineen vaihtoehtoiseksi väitetyillä *saper-* ja *sapere*-infinitiivimuodoilla on osoittautunut olevan omat syntagmaattiset valintansa. Näistä ensimmäistä käytetään nimittäin yleensä infinitiivimuotojen, jälkimmäistä sivulauseiden yhteydessä. Suomen kielen *kauniiksi* ja *kauniilta* saavat puolestaan selvästi erilaiset kollokaatit ja semanttiset preferenssit: *kauniiksi* <mm. SANOJA, KEHUA, TEHDÄ, MUUTTUA> (eli puhumista ja muuttamista ilmaisevat kollokaatit) ja *kauniilta* <NÄYTTÄÄ, VAIKUTTAA, KUULOSTAA> (eli näyttää tai kuulostaa joltakin) (Jantunen 2004a: 27–28). Vastaavanlaisia oletuksia eri taivutusmuotojen kontekstuaalisesta sidoksisuudesta on muitakin (ks. Sinclair 1991: 8; Aston & Burnard 1998: 8). Sinclair (1991: 8) ja Stubbs (2001:17) jakavatkin käsityksen, että leksikaalisia (fraseologisia) yksiköitä pitäisi tarkastella lemmatasolla ainoastaan, mikäli lemman taivutusmuotojen syntagmaattiset valinnat olisivat yhteneviä. Tämä toteutunee kuitenkin harvoin.

Sen lisäksi että kielenoppijalla on runsaasti mahdollisuuksia päätyä epätyypillisiin leksikaalis-semanttisiin rakenteisiin tuottaessaan kohdekielen fraseologisia yksiköitä (ks. Jantunen 2009b), ongelmia tuottavat myös morfologiset piirteet nimenomaan fraseologian näkökulmasta.

Monet yksittäiset morfofonologian ja morfosyntaksin piirteet liittyvätkin fraseologisten sääntöjen tunnistamisen, osaamisen ja opettamisen vaikeuteen: oppijalla ei ole yleensä tietoa ilmausten frekventeimmistä ja siten myös osaamisen kannalta tärkeimmistä muodoista eikä myöskään niiden kotekstuaalisista (kuten kolligationaalisista) ja tekstinulkoisen maailman (tilannekontekstin) suhteista. Esimerkiksi valenssisääntöjen hallinta pohjautuu fraseologiasääntöjen havaitsemiseen ja ymmärtämiseen ja sisältää niin leksikaalisten, kieliopillisten ja semanttisten samoin kuin tekstin ulkoisten säännösten osaamisen. Tämä monimutkainen sääntöverkosto on natiivin kielenkäyttäjän pääasiassa implisiittisesti hallitsema, eikä sen auki purkaminen ole yksiselitteistä kokeneellekaan opettajalle.

5. Aineisto ja menetelmät

Tapaustutkimusten aineistona käytämme Oulun Kansainvälistä oppijansuomen korpusta (ICLFI, ks. Jantunen 2011b). Aineisto koostuu ulkomaisissa yliopistoissa opiskelevien suomenoppijoiden kirjoitetuista teksteistä. Korpuksen tekstit ovat opetustilanteessa tuotettuja harjoitustöitä tai kokeita. Aineistossa on sekä fiktiivisiä (mm. kertomuksia ja kirjeitä) että asiatekstejä (mm. arvosteluja ja esseitä), ja niiden kokonaisuudessa muodostaa maaliskuussa 2012 n. 5 300 tekstin ja n. miljoonan saneen tekstikokonaisuuden.

Tätä tutkimusta varten on aineistosta muodostettu kaksi osakorpusta: ICLFI-viro on osakorpus, joka sisältää noin 104 000 saneen (579 tekstiä) verran vironkielisten oppijoiden kirjoituksia. ICLFI-multi on puolestaan niin sanottu moniäidinkielen korpus, jossa on tekstejä eri kieliä äidinkielenään puhuvilta opiskelijoilta. Suurimmat aineistot muodostuvat puolaa, venäjää, saksaa, ruotsia, tšekkiä, kiinaa ja hollantia äidinkielenään puhuvien suomenoppijoiden teksteistä. Yhteensä eri äidinkielifaustoja on ICLFI-multissa 21, jokin eräiden osakorpusten tekstit (kuten ranskan-, japanin- ja marinkielisten tekstit) muodostavat hyvin pienen tekstijoukon. ICLFI-multi-aineiston koko on noin 698 000

sanetta ja 3 144 tekstiä. Vertailuaineistona käytetään Käännössuomen korpuksesta (Mauranen 2000) luotua natiivisuomen aineistoa, joka on noin 4 miljoonan saneen fiktiivisiä ja asiatekstejä sisältävä korpus.

Kahden oppijansuomen tutkimusaineiston käyttöä motivoi kolmivaiheinen vertailumenetelmä (engl. *Three-Phase Comparative Analysis, TPCA*), jota on aiemmin testattu käännöskielen tutkimuksessa (Jantunen 2004b; ks. myös Kubáčková 2009; Dose 2010). Menetelmässä yksi osakorpus muodostuu kielitaustaltaan vain yhden yhtenäisen ryhmän teksteistä (tässä vironkielisten tekstit), kun taas toinen osakorpus koostuu kielitaustaltaan varioivista teksteistä eli se edustaa tutkittavaa varianttia yleensä, tässä tutkimuksessa siis laajasti oppijansuomea. Kolmivaiheinen vertailu syntyy tarkasteltaessa oppijankieliaineistoja suhteessa toisiinsa ja molempia erikseen suhteessa natiiviaiaineistoon. Vertailun tavoitteena on saada selville, poikkeavatko aineistot toisistaan joidenkin kielenpiirteiden osalta. Jos eroja esiintyy korpusten välillä, voidaan olettaa, että erot johtuvat kielitaustasta. Kuitenkin kuten todellisia tekstejä käyttävissä tutkimuksissa yleensä, on myös nyt pidettävä mielessä se, että monien muiden taustamuuttujien (kuten muiden kielten osaamisen ja vaikkapa oppimateriaalin merkityksen) poissulkeminen on mahdotonta.

Tässä artikkelissa tutkittavien kielenainesten valinta on tehty korpusvetoisesti. Korpusvetoisuus tarkoittaa yksinkertaistetusti sitä, että tutkittavia aineksia ei valita aiempiin tutkimuksiin perustuen tai tutkijan intuition perusteella, vaan ne nousevat aineistosta esiintymistiheytensä perusteella. Yksi tavallisimmista korpusvetoisista menetelmistä on avainsana-analyysi. Siinä aineistoja verrataan tilastollisin menetelmin keskenään, ja vertailun tulokseksi saadaan avainsanoja eli sanoja (tai sananmuotoja), jotka esiintyvät tutkimusaineistossa huomattavasti tiheämmin kuin vertailuaineistona käytettävässä materiaalissa. (Korpusvetoisuudesta ja avainsana-analyysistä tarkemmin esim. Scott & Tribble 2006; Jantunen 2009a; Jantunen 2012.) Tutkimuksessamme olemme käyttäneet lähtökohtana koko ICLFI:n vertailua natiiviaiaineistoon saadaksemme selville, millaiset tutkimusaineiston sanat poikkeaa-

vat frekvenssiltään natiiviaineistosta. *WordSmith Toolsin* (Scott 2004) *Keywords*-työkalun avulla saatu avainsanalista, eli oppijoiden suosimien sanojen lista, sisältää ensinnäkin runsaasti sellaisia sanoja, joiden runsaan lukumäärän voi selittää opiskelijoiden kirjoittamien tekstien tehtävänäannoilla: sanat kuten KIINA, SINUHE, TIMO, YLIOPISTO ja OPISKELLA selittyvät helposti kirjoitustehtävillä eivätkä siis ole mielenkiintoisia, kun tavoitteena on kuvata nimenomaan kielenoppisen ja oppijankielen piirteitä. Tutkimukseemme olemme valinneet avainsanalistalta sanat IHMINEN ja PITÄÄ, jotka esiintyvät eri muodoissaan tilastollisesti merkittävimpien oppijankielen avainsanojen listalla, kun oppijanaineistoa verrataan natiiviaineistoon; IHMINEN-nominin tilastollisesti yliedustunein muoto on *ihmiset* ja PITÄÄ-verbien *pitää* ja *pidän*.

Seuraavaksi tarkastelemme, miksi tietyt muodot ja koko lemma ylipäättäänkin esiintyvät oppijankieliaineistossa niin taajan. Tavoitteemme on etsiä morfologian ja fraseologian yhtymäkohtia ja tarkastella fraseologisten piirteiden selitysvoimaa morfologisten piirteiden epätyypillisten esiintymämäärien kuvauksessa. On toki luonnollista, että kahden sanan tarkastelulla ei voida laajasti selittää oppijansuomen morfologista primingia ja fraseologisten piirteiden syy-yhteyttä morfologisessa primingissa onnistumiseen. Tuloksia on siten pidettävä alustavina ja tutkimusta pilottiluonteisena: tarkoituksena on selvittää morfologisen primingin olemusta, hypoteeseja ja lisäksi myös aineiston merkitystä tutkimuksessa.

6. Tapaustutkimukset IHMINEN ja PITÄÄ

6.1. Tapaus IHMINEN

IHMINEEN-nominin analyysissä keskitymme kieliopillisten sijojen analyysiin, sillä ne ovat odotetusti taajimmin esiintyviä muotoja aineistoissa ja näin ollen analyysin näkökulmasta myös frekvenssiltään riittäviä (kieliopillisten sijojen osuus kaikista sijoista on jokaisessa aineistoissa yli 80 %). Kun kieliopillisissa sijoissa taivutettujen muotojen ja muiden muotojen yhteenlaskettua määrää tarkastellaan tilastollisesti, paljas-

tuu alustavasti, että muotojen määrät poikkeavat aineistoittain erittäin merkitsevästi toisistaan ($\chi^2 = 256,8$ (df = 6, $p < 0,0001$, kriittinen arvo 22,46)). Näyttää siis siltä, että oppijan- ja natiivikielen aineistot erottuvat toisistaan siinä, missä kieliopillisissa sijoissa IHMINEN esiintyy näissä kielivarianteissa. Yksityiskohtaisempi tarkastelu paljastaa, että molemmissa oppijanaineistoissa kieliopillisissa sijoissa taivutettujen muotojen määrä poikkeaa natiiviaineistosta ja että ICLFI-multi-aineisto poikkeaa natiivituotoksista enemmän kuin vironkielisten oppijoiden kirjoittamat tekstit. Sekä yksikkö- että monikkomuotojen yli- ja alikäyttö on kirjattu muodoittain ja aineistoittain taulukkoon 1 tilastollisesti merkitsevimmin erottuvien muotojen osalta. Moniäidinkielisessä aineistossa partitiivin yksikkömuotoa lukuun ottamatta kaikkien kieliopillisissa sijoissa esiintyvien muotojen taajuudet poikkeavat natiivikielen aineistosta. Sen sijaan virolaisten suomenoppijoiden tuotokset noudattavat sijojen osalta natiivin kaltaista kielenkäyttöä, poikkeuksena jälleen partitiivin yksikkömuodot, joita vironkieliset oppijat käyttävät runsaammin kuin natiiviaineisto antaisi olettaa.

Seuraavaksi keskitymme muotojen ylikäyttötapauksiin². Muotojen tavanomaista suuremmille frekvensseille tarjoutuu useita selityksiä, joiden juuret palautuvat kielen fraseologialuonteeseen. Ensinnäkin NOM. PL-muodon ylikäyttö ICLFI-multi-aineistossa selittyy osittain impersonaalisen passiivin (esim. *mennään*) välttämällä ja sen korvaamisella *ihmiset*-rakenteella (1, 2). Geneerinen IHMINEN passiivin korvaajana (mm. nollapersoonan ohella) onkin Seilosén (2011) mukaan yksi suomenoppijoiden strategisista keinoista; impersonaalisen passiivin onnistunut käyttö puolestaan lisääntyy kielitaidon kehittyessä (Martin ym. 2010)³.

² Myös alikäyttötapaukset ovat oppijankieliaineiston kuvauksen kannalta relevantteja, mutta käsillä olevassa tutkimuksessa ei ole mahdollista perehtyä niihin, koska sellaisten elementtien analyysi, joita aineistossa ei ole, vaatii erityisen tarkan perehtymisen natiivikieliseen tuotokseen ja tulosten vertailun muihin aineistoihin. Tässä tutkimuksessa on kuitenkin otettu lähtökohdaksi oppijankieliaineistot.

³ Passiivin lisääntyminen näkyy myös ICLFI-aineistossa, sillä passiivimuotojen määrät moninkertaistuvat siirryttäessä alkeista keskitasolle ja edistyneelle tasolle.

TAULUKKO 1. *Kieliopillisten sijojen yli- ja alikäyttö oppijankielen aineistoissa natiiviaineistoon verrattuna*

(*** = erittäin merkitsevä, $z > 3,29$ (= $p < 0,001$))⁴

<i>ihminen-nomini</i>		ICFLI-viro	ICFLI-multi
NOM.SG	<i>ihminen</i>	–	alikäyttö***
NOM.PL	<i>ihmiset</i>	–	ylikäyttö***
GEN.SG	<i>ihmisen</i>	–	alikäyttö***
GEN.PL	<i>ihmisten/ihmisten</i>	–	alikäyttö***
PTV.SG	<i>ihmistä</i>	ylikäyttö***	–
PTV.PL	<i>ihmisiä</i>	–	ylikäyttö***

- (1) Kirkon jälkeen **ihmiset** menevät kotiin syömään. (SA0106A)⁵
- (2) --- on kuvaavaa, että **ihmiset** puhuvat murteita vain pienemmissä kylissä eikä kaupungeissa. (UN0014A)

Tämän lisäksi oppijat käyttävät toistuvasti kyseistä NOM.PL-muotoa eksplisiittisissä ja redundanteissa rakenteissa, joissa sitä edeltää kvanttoripronomini (3 ja 4). Kvanttoripronominin, kuten KAIKKI, MONI, JOKAINEN ja JOKU, viittaavat yksinäänkin esiintyessään joukkoon ihmisiä (*Monelle on varmaan tuttu tilanne se, että ---*, ISK 2004 § 744), joten

⁴ NOM.SG: ICFLI-viro vs. natiiviaineisto $z = 0,348$ ($p > 0,05$), ICFLI-multi vs. natiiviaineisto $z = 8,059$ ($p < 0,001$)

NOM.PL: ICFLI-viro vs. natiiviaineisto $z = 0,722$ ($p > 0,05$), ICFLI-multi vs. natiiviaineisto $z = 15,417$ ($p < 0,001$)

GEN.SG: ICFLI-viro vs. natiiviaineisto $z = 0,873$ ($p > 0,05$), ICFLI-multi vs. natiiviaineisto $z = 15,884$ ($p < 0,001$).

GEN.PL: ICFLI-viro vs. natiiviaineisto $z = 1,617$ ($p > 0,05$), ICFLI-multi vs. natiiviaineisto $z = 4,391$ ($p < 0,001$).

PTV.SG: ICFLI-viro vs. natiiviaineisto $z = 10,941$ ($p < 0,001$), ICFLI-multi vs. natiiviaineisto $z = 0,492$ ($p > 0,05$).

PTV.PL: ICFLI-viro vs. natiiviaineisto $z = 0,589$ ($p > 0,05$), ICFLI-multi vs. natiiviaineisto $z = 7,611$ ($p < 0,001$).

⁵ Koodissa kirjainyhdistelmä yksilöi äidinkielen (ES = espanja, IS = islanti, KI = kiina, PU = puola, SA = saksa, UN = unkari, VE = venäjä, VI = viro), numerosarja kirjoittajan ja viimeinen merkki tekstin.

NOM.PL-muodon käyttö ei sinänsä tuo ilmaukseen lisäinformaatiota vaan on merkityksen näkökulmasta redundantti. Fraseologisesta näkökulmasta tarkasteltuna *ihmiset* näyttääkin kolligoivan ICLFI-multi-aineistossa epätyypillisen taajaan kvanttoripronominien kanssa: kun natiiviaineistossa *kaikki ihmiset* ja *monet ihmiset* esiintyvät miljoonaa sanetta kohti kuusi ja seitsemän kertaa, ovat multi-aineiston vastaavat suhdeluvut 54 ja 131.

- (3) Yhtäkkiä kaikki **ihmiset** huutaivat ja matkatavaroita sijaitsivat lattialla. (SA0023B)
- (4) Mutta monet **ihmiset** sanovat Pekingin ilmanlaatu ei ole hyvä ---. (KI0016Q)

Toinen samassa aineistossa yliedustunut muoto on PTV.PL *ihmisiä*. Taa-
jan käytön osittaiseksi selitykseksi paljastuvat eksistentiaalilauseet, joissa kvantifioidaan ihmisten määrää jossakin tietyssä paikassa (5). Muoto on odotuksenmukainen kvanttoriadverbien jälkeen, sillä niitä seuraa NP.PTV (ISK 2004 § 657). Yliedustumisen syynä ei siten ole muodon käyttö sinänsä vaan pikemminkin se, että oppijoilla on kirjoituksissaan (kuten esimerkiksi kuvauksissa) taipumus korostaa vastaanottajan kannalta tai omasta mielestään tärkeitä, uniikkeja tai outoja ilmiöitä (Jantunen 2007; Kallioranta 2009). Oppijat korostavat teksteissään ihmispaljouden lisäksi usein myös oletettuun tai todelliseen kokemuspiiriin kuulumattomia ilmiöitä käyttämällä eksistentiaalilauseissa juuri kvantifointia (6). *Paljon*-kvantifointi onkin erittäin yleistä suomenoppijoiden teksteissä, ja *paljon* on myös yksi niin sanotuista oppijansuomen avainsanoista (Jantunen 2011a).

- (5) Minun täytyy sanoa, että Puolassa on paljon viisaita **ihmisiä**. (PU0008D)
- (6) Lapissa on paljon lunta. (KI0010M)

Muodon runsaudelle on lisäksi löydettävissä selitys kvanttoriadverbien ja kvanttoripronominin sekoittumisesta: jopa 29 % *paljon ihmisiä* -tapauksista olisi korvattava kvanttoripronominin sisältävällä rakenteella (esim.

monet ihmiset tai *useat ihmiset*), jolloin myös nominin pitäisi olla odotuksenmukaisesti toisessa sijassa (NOM.PL), kuten esimerkissä (7), johon sopisi paremmin *monet/useat ihmiset*.

- (7) Selitetäänkin, että paljon **ihmisiä** jatkavat opiskella aikuisina, jopa työn ohessa. (IS0002I)

Tässä yhteydessä on kuitenkin huomattava, että oppijat ovat hallinneet NP:tä koskevan sijamuotovalinnan (*paljon* + NP.PTV), mutta tuottavat odotuksenvastaisen muodon nimenomaan kvanttorin valinnan näkökulmasta. Kun PTV.PL-ylikäyttötapauksia tarkastelee jälleen fraseologisuuden näkökulmasta, voi havaita selvästi, että ICLFI-multi-aineistossa PTV.PL kolligoi siis voimakkaasti kvanttoreiden ja eksistentiaalilauseiden kanssa, ja erityisesti kollokointi kvanttoriadverbin *paljon* kanssa on tyyppillistä.

Kolmas ylikäyttötapaus löytyy edellisistä poiketen vironkielisten oppijoiden tuotoksista: *ihmistä* (PTV.SG) esiintyy tässä osakorpuksessa merkitsevästi useammin kuin yleensä kohdekielisessä tuotoksessa. Yhdeksi selitykseksi nousevat aineistosta jälleen eksistentiaalirakenteet, tällä kertaa esittelevät, perhettä kuvaavat rakenteet: *perheessä(ni) on # ihmistä* (8 ja 9). Rakenteita on lukumääräisesti vähän ($n = 9$), mutta ne kattavat PTV.SG-tapauksista kuitenkin kolmasosan. (Rakenne näyttää olevan tätäkin taajemmin esiintyvä, sillä rakenteen määrä nousee 63:een, kun aineistoa laajennetaan tutkimusaineiston kokoamisen jälkeen saaduilla teksteillä; PTV.SG-muodossa oleva ilmaus voi olla myös (*perheen*)jäsentä.) Näissä rakenteissa PTV.SG määrittää rakenteen edus-sanana olevaa nominatiivimuotoista numeraalia ja on tässä rakenteessa siis odotuksenmukainen. Näille rakenteille löytyy selitys opetuksessa annetuista tehtävänannoista, sillä ovathan minun perheeni -tyyppiset tekstit erittäin tavallisia juuri alkeistason opetuksessa, kun harjoitellaan sukulaisuusnimityksiä ja esimerkiksi lukusanoja, omistus- ja eksistentiaalirakenteita. Näin ollen ylikäytön syyksi paljastuu ainakin osittain itse tehtävänanto eikä kysymyksessä ole tältä osin niinkään kieleen liittyvä ilmiö.

- (8) Minun perheessä on yhdeksän **ihmistä**. Perheeseeni kuuluu äiti, isä, 3 siskoa... (VI0284)
- (9) Minun perheessäni on viisi **ihmistä** ja minä on olen nuorin perhejäsen. (VI0064E)

Ymmärrettävyyden ja kieliopillisuuden näkökulmasta tässä *perheessäni on # ihmistä* -rakenteessa ei ole mitään ongelmaa: kirjoittajat noudattavat suomen kielen tyypillistä eksistentiaalirakenteen kaavaa, ja ilmaus on läpinäkyvä ja ymmärrettävä. Yllättäen natiiviaineistoja analysoimalla kuitenkin paljastuu, että rakenne ei ole tavallinen suomalaisissa kirjoituksissa teksteissä. Tutkimuksessa käytetyssä natiiviaineistossa rakennetta ei esiinny ollenkaan, mikä voi tosin johtua tekstien aihepiireistä. Lisävarmistusta varten tehty Google-hakukoneella suoritettu haku⁶ vahvistaa rakenteen kuitenkin erittäin harvinaiseksi: sitä löytyy internetistä hakukoneen avulla vain kolmetoista tapausta. Näin ollen rakenne paljastuu käytön kannalta epätyypilliseksi; selvästi tavallisempi tapa kertoa perheenjäsenistä on luetteloida henkilöt hyponyymeinä ilman merkitykseltään yleistä *ihmistä*-sanaa, esimerkiksi *Perheessäni on isä, äiti ja 2 siskoa*. Tekstien taustatietojen analyysi paljastaa mielenkiintoisesti, että *perheessäni on # ihmistä* -rakenteet ovat peräisin yhdeltä opiskelijaryhmältä, nimittäin tarttolaisilta suomen opiskelijoilta (edellä mainituista 63 tapauksesta vain kolme tekee tästä poikkeuksen). Rakenteen käyttöön ei kuitenkaan löydy selitystä oppikirjasta, sillä siinä rakennetta ei ole esitelty eivätkä samaa oppikirjaa käyttävät tallinnalaiset opiskelijatkaan ole tuottaneet näitä rakenteita. On myös hyvä huomata, että kun nämä erityiset rakenteet poistetaan aineistosta, PTV.SG on edelleen tilastollisin perustein ylikäytetty. Tiivistäen voi esittää, että muoto PTV.SG kolligoi eksistentiaalilauseen ja numeraalien kanssa ja sillä on ainakin osittainen semanttinen preferenssi 'perhe' (etenkin tarttolaisilla opiskelijoilla).

⁶ Hakulausekkeet "*perheessäni on * ihmistä*" / "*perheessä on * ihmistä*". (Google-hakukoneella on mahdollista hakea tietoa myös korvausmerkein, tässä asteriskin avulla.)

6.2. Tapaus PITÄÄ

Toisena tapauksena tutkimuksessamme on monimerkityksinen verbi PITÄÄ. Verbin täydennykset ja eri merkitykset ovat yhteydessä toisiinsa (ISK 2004 § 449), joten kun tarkastellaan verbin kerasanojen morfologisia piirteitä, ei voida välttää tutkimasta myös verbin eri merkityksiä. “Kielitoimiston sanakirja” (KS s.v. *pitää*) antaa verbille 18 erilaista merkitystä, joista tärkeimmät voidaan kuitenkin niputtaa neljän erilaisen rakenteen alle. Ensimmäinen on transittiivisen PITÄÄ-verbin ja sen olosijaisen täydennyksen muodostama rakenne, jota “Iso suomen kielioppi” (2004 § 475) nimittää pitämismuotiksi. Rakenteen adverbiaalitäydennys voi olla lokatiivinen, habitiivinen tai predikoiva. (*Pitää jotakin kädessä ~ pystyssä ~ käymässä.*) Tämän rakenteen alle niputtuvat verbin jonkinlaiseen säilyttämiseen, pitämiseen ja pysyttämiseen (otteessa, asemassa, tilassa) liittyvät merkitykset (10).

Toinen selvästi erottuva merkitysrypäs liittyy kannanoton ja mielihiteen ilmaisemiseen. Siihen liittyvät merkitykset jakaantuvat rakenteellisesti kahteen ryhmään: Jonkin arvioimista jonkinlaiseksi (‘pitää jotakin jonakin’) “Iso suomen kielioppi” (2004 § 485) nimittää arviointimuotiksi (11). Siinä PITÄÄ-verbiin liittyy essiivimuotoinen predikatiiviadverbiaali ja (partitiivi)objekti, joka voi olla myös A-infinitiivi. Toinen arviointiin liittyvä rakenne ilmaisee ‘tykkäämistä’, ja siinä verbi saa rektiokseen elatiivimuotoisen täydennyksen (12). PITÄÄ-verbi esiintyy myös nesessiivirakenteessa (13), jolloin se on yksipersonainen: sen predikaattiverbi on aina yksikön kolmannessa persoonassa. Rakenteen ilmisubjekti on genetiivimuotoinen, ja subjektittomissa tapauksissa lause tulkitaan yleensä nollapersonaiseksi (ISK 2004 § 1354). PITÄÄ-verbillä on runsaasti myös muita merkityksiä, joista useille on tyypillistä fraasimaisuus (esim. PITÄÄ LOMAA, PITÄÄ HAUSKAA, PITÄÄ HUOLTA).

- (10) Hän **pitää** kädessään sateenvarjoa.
- (11) Hän **pitää** tätä hyvänä ideana.
- (12) Hän **pitää** suklaasta.
- (13) Hänen **pitää** lähteä.

Kielenoppijoiden tuottamien verbinmuotojen frekvenssit ja variaatiot eroavat sekä osakorpuksittain toisiinsa että suhteessa natiivisuomeen. Muotojen kirjo on odotetusti kapeampi kuin natiivipuhujien, ja oppijansuomessa on odotuksenmukaisesti enemmän virheellisiä muotoja kuin toimitettuun aineistoon pohjautuvassa natiivikielen vertailuaineistossa, jossa ainoat norminvastaiset muodot ovat murteellisia *pittää*-muotoja. Mielenkiintoista on, että ICLFI-multi-aineisto eroaa sekä muotojen variaatiota että frekvenssejä tarkasteltaessa vähemmän natiivikielestä kuin ICLFI-viro, sillä sukukielisten opiskelijoiden tuotosten voisi odottaa olevan lähempänä kohdekieltä (Ks. Lähdekielen vaikutuksesta Kaivapalu 2005; Kaivapalu & Eslon 2011). Edellä tarkastellussa IHMINEN-tapauksessahan tilanne oli juuri näin.

Erityisen kiinnostavia ovat oppijankielessä hyvin yleiset muodot *pidän* ja *pitää*, joiden esiintymien määrät eroavat sekä ICLFI-multi- että ICLFI-viro-aineistossa merkitsevästi natiivianeistosta⁷. Merkitsevin ero on muodossa *pidän*, joka on ICLFI-viro-aineiston yleisin muoto ja kattaa 36 prosenttia kaikista PITÄÄ-verbin muodoista. ICLFI-multi-aineistossakin se on hyvin yleinen (24 %), kun taas natiivisuomessa sitä käytetään vain hyvin vähän (1,3 %)⁸. Tämä osoittaa, että vaikka ICLFI-multi-aineiston fraseologisen yksikön ytimessä olevan PITÄÄ-verbin morfologinen priming on selvästi erilainen kuin natiivikielessä, eroaa virolaisten kielenoppijoiden priming sitäkin enemmän kohdekielestä.

Seuraavaksi siirrymme ytimen morfologisesta primingista sen kerasanojen morfofonologisiin muotoihin ja tarkastelemme erilaisia täydennysympäristöjä, jotka reaalistuvat verbin eri merkityksissä. Merkitysvarianttien käyttö eroaa jälleen sekä osakorpuksittain että suhteessa natiivisuomeen. Analyysin kohteeksi on valikoitunut ytimen

⁷ ICFLI-viro vs. natiivianeisto $\chi^2 = 2113,99$, ICLFI-multi vs. natiivianeisto $\chi^2 = 2049,10$; ICLFI-viro vs. ICLFI-multi $\chi^2 = 51,16$ (df = 2, p < 0,0001, kriittinen arvo 13,82).

⁸ Esiintymien absoluuttiset määrät ovat ICLFI-virossa 182/505, ICLFI-multissa 732/3 103 ja natiivisuomessa 135/10 439 (*pidän*-muoto/kokonaismäärä).

morfologisessa primingissa esille nouseen, oppijankielessä yleisen *pidän*-muodon merkitysten jakautuma aineistoissa. Vaikka pitelymuotin kuvaama merkitys ‘pitää jotakin kädessään’ on sanakirjoissa (KS s.v. *pitää*) annettu yhdeksi perusmerkityksistä, jää se analyysin ulkopuolelle, koska se on sekä natiivisuomen että oppijankielen aineistossa hyvin harvinainen. Lisäksi nesessiivinen merkitys ‘täytyä’ on kieliopillisesti mahdoton ensimmäisen persoonan muodossa *pidän*. Näin ollen vertailu natiivikäyttöön ei ole mielekäästä, vaikka suomenoppijat ovatkin tuottaneet jonkin verran nesessiivisiä merkityksiä ja rakenteita yksikön 1. persoonan muodosta. Ne näkyvät myöhemmin taulukossa 3. Muut aikaisemmin mainitut perusmerkitykset jakautuvat taulukon 2 kuvaamalla tavalla.

TAULUKKO 2. *Pidän-muodon ali- ja ylikäyttö oppijankielen aineistoissa natiivisuomeen verrattuna*

(*** = erittäin merkitsevä, $z > 3,29$ (= $p < 0,001$))⁹

<i>Pidän</i>-muodon merkitykset	ICFLI-viro	ICFLI-multi
‘tykätä’	ylikäyttö***	ylikäyttö***
‘arvioida’	alikäyttö***	alikäyttö***
fraasit	alikäyttö***	alikäyttö***

Sekä ICLFI-viro- että ICLFI-multi-aineistoissa merkitys ‘tykätä’ on selvästi ylliedustunut ja essiivimuodon sisältävä arviointimuotti (‘pitää jotakin jonakin’) on aliedustunut. Itse asiassa virolaiset suomenoppijat käyttävät liki yksinomaan *pidän*-muotoa merkityksessä ‘tykätä’: sen osuus on 99 prosenttia kaikista muodon esiintymistä (14, 15). Elatii-

⁹ ‘tykätä’: ICFLI-viro vs. natiiviaineisto $z = 11,488$ ($p < 0,001$), ICLFI-multi vs. natiiviaineisto $z = 13,606$ ($p < 0,001$)

‘arvioida’: ICFLI-viro vs. natiiviaineisto $z = 6,560$ ($p < 0,001$), ICLFI-multi vs. natiiviaineisto $z = 6,130$ ($p < 0,001$)

Fraasit: ICFLI-viro vs. natiiviaineisto $z = 8,840$ ($p < 0,001$), ICLFI-multi vs. natiiviaineisto $z = 13,677$ ($p < 0,001$).

vitäydennyksisen ‘tykätä’-merkityksen prosenttiosuus on suuri myös ICLFI-multi-oppijoilla (90 %), mutta tässä aineistossa esiintyy jonkin verran (6 %) myös essiivirakenteita (16), vaikkakin ero kohdekieleen on selvä: natiivisuomessa ‘tykätä’-merkityksiä on 42 prosenttia ja essiivirakenteellisia arviointeja 22 prosenttia kaikista käyttötavoista. Myös fraasimaisia ilmauksia oppijansuomessa käytetään selvästi vähemmän kuin natiivisuomessa. Virolaiset eivät tuota niitä lainkaan, ja ICLFI-multi-aineiston suomenoppijoilla niitä on 2 prosenttia kaikista *pidän*-muodoista (17). Natiivisuomessa fraasien osuus on niinkin suuri kuin 36 prosenttia. Tämä osoittaa, että myös verbin täydennyksenä olevien kersanojen morfologiset primingit eroavat selvästi kohdekielestä ja myös tässä tapauksessa ICLFI-viro eroaa siitä enemmän kuin moniäidinkielen aineisto.

- (14) **Pidän** kovin musiikista. (VI0276C)
- (15) **Pidän** tämästä elokuvasta. (ES0007)
- (16) **Pidän** suomea vaikeana kielenä. (SA0024)
- (17) **Pidän** hauskaa aina. (KI0009F)

Seuraavaksi tarkastellaan tarkemmin oppijakielelle tyypillisen *pidän*-muodon yllättävän vahvaa semanttista primingia merkityksen ‘tykätä’ kanssa. Taulukko 3 esittää, miten *pidän*-muodot jakautuvat taitotasoitain arvioituina.

Liki kaikki viron aineiston *pidän*-muodot ovat alkeistasolla olevien opiskelijoiden tuottamia. Se, että merkitys ‘tykätä’ on korostunut (99 %), selittyy ainakin osin sillä, että rakennetta *pitää jostakin* harjoitteluun kielenoppimisen alkuvaiheessa runsaasti, mutta se ei kuitenkaan selitä kyseisen fraseologisen yksikön täydellistä puuttumista muilta taitotasoilta. Toinen mahdollinen selitys on, että ICLFI-viron aineistossa yleinen tehtävätyyppi, kuvaus, ja tehtävänannot (*Perheeni, Päiväni*) ohjaavat käyttämään kyseistä rakennetta. Jälkimmäinen vaihtoehto indikoi sitä, että rakenteen ja sen myötä ‘tykätä’-merkityksen ja tekstilajien sekä myös tehtävänantojen välillä on Hoeyn (2005; 2007) mainitsemaa tekstuaalista primingia.

TAULUKKO 3. *Pidän-muodon jakautuminen eri merkityksiin taitotasoittain oppijankielen aineistoissa ja natiivisuomessa.*

(Alkeistaso: < 200 tuntia opetusta; keskitaso: 200–400 tuntia opetusta; edistyneet: > 400 tuntia opetusta)

<i>Pidän-</i> <i>muodon</i> <i>merki-</i> <i>tykset</i>	ICLFI-viro			ICLFI-multi			natiivi- suomi
	alkeis- taso	keski- taso	edisty- neet	alkeis- taso	keski- taso	edisty- neet	
‘tykätä’	177 (99 %)	2	–	351 (94 %)	203 (92 %)	107 (76 %)	57 (42 %)
‘arvioida’	0 %	–	1	0 %	13 (6 %)	29 (21 %)	29 (22 %)
fraasit	0 %	–	–	9 (2 %)	4 (2 %)	4 (3 %)	49 (36 %)
‘täytyä’	2 (1 %)	–	–	6 (2 %)	0 %	0 %	0 %
muut	0 %	–	–	6 (2 %)	0 %	0 %	0 %
yhteensä	179 (100 %)	2	1	372 (100 %)	220 (100 %)	140 (100 %)	135 (100 %)

Analyysi ei paljasta, onko rakenne yhtä dominoiva ICLFI-viro-aineistossa muidenkin PITÄÄ-verbin muotojen kohdalla, eikä sitä, onko muiden muotojen merkitys jakauma ehkä toisenlainen. Kovin merkittäviä eroja ei kuitenkaan löytyne, koska aiemminhan kävi jo ilmi, että nimenomaan muoto *pidän* on viron aineistossa selvästi yleisin ja muiden muotojen käyttö on hyvin vähäistä. ICLFI-multi-korpuksen analyysi kertoo, että myös tässä aineistossa alkeis- ja keskitasolla merkitys ‘tykätä’ on selvästi yleisin, molemmissa yli 90 prosenttia. Edistyneillä prosenttiosuus (76 %) on kuitenkin jo selvästi pienempi, joten tämän aineiston perusteella voimme sanoa, että verbimuodon käyttö lähestyy vähitellen natiivikäyttöä (42 %) taitotason parantuessa. Selvimmin tendenssi näkyy kuitenkin tarkasteltaessa essiivilaajennuksen sisältävää merkitystä ‘arvioida’. Näitä rakenteita edistyneet ICLFI-multi-oppijat käyttävät liki yhtä usein kuin natiivipuhujat. Prosenttiosuuksissa on vain yhden prosenttiyksikön ero (21 % ja 22 %). Rakenteet löytyvät usein esseistä ja mielipidekirjoi-

tuksista, joiden kanssa muodolla ja siihen liittyvällä arviointimerkityksellä tuntuisi olevan tekstuaalista primingia.

On kuitenkin pantava merkille, että verbin fraasimainen käyttö ei lisääny liki lainkaan taitotason kasvaessa. Edistyneilläkin kielenoppijoilla *pidän*-muodon sisältäviä fraaseja on vain 3 prosenttia, kun vastaava luku natiivisuomessa on peräti 36 prosenttia. Mielenkiintoisella tavalla – tämän analyysin perusteella – ilmiö tiivistyy kuitenkin juuri *pidän*-muotoon, sillä analysoitaessa *pitää*-muotoa (joka voi olla paitsi 1INF myös 3SG) nähdään, että prosenttiosuudet ovat liki samat: ICLFI-multi-aineistossa edistyneiden kielenoppijoiden tuottamista *pitää*-muodoista on fraasimaisia ilmauksia 19 prosenttia ja natiivisuomessa 20 prosenttia. Voimmekin sanoa, että fraaseilla (*PITÄÄ LOMAA*, *PITÄÄ HAUSKAA*, *PITÄÄ HUOLTA*) on oppijankielessä positiivinen morfologinen priming muotoon *pitää*, mutta negatiivinen priming muotoon *pidän*. Lisäksi ICLFI-multi-aineiston *pitää*-muodon fraasien frekventuaalinen osuus on hyvin samanlaista kuin natiivisuomessa; siltä osin morfologinen priming on hyvin kohdekielimäistä, päinvastoin kuin *pidän*-muodossa. Mahdollinen selitys ilmiölle on, että fraasit opitaan usein juuri perusmuodoissaan. Vaikka oppijakielen fraasit ovatkin usein yksinkertaisia fraseologisia yksiköitä kuten *pitää huolta*, *pitää hauskaa* (18), ei fraasien käyttö rajoitu vain niihin vaan fraasien joukossa on mukana myös muita, esimerkiksi esittämiseen ja järjestämiseen liittyviä fraaseja (19, 20).

(18) Ei ole paljon aikaa **pitää** hauskaa. (KI0014D)

(19) --- psykologi **pitää** esitelmän itsemurhasta ---. (PU0005O)

(20) Faarao **pitää** suuret juhlat maansa kunniaksi. (VE00100)

7. Morfofonologiset virheet sanan taivutuksessa

Yksi hypoteeseistamme oli, että äidinkielenään sukukieltä puhuvat kielenoppijat (tässä tutkimuksessa virolaiset) tuottavat morfofonologisesti epätyypillisiä ja virheellisiä taivutusmuotoja vähemmän kuin muut, kielitaustaltaan esimerkiksi indoeurooppalaiset opiskelijat. Hypoteesin

testaamiseksi laskimme kaikki IHMINEN-nominin tuotoksissa esiintyvät epätyypillisyydet. Taivutusmuodoissa esiintyi virheitä, jotka on tulkittavissa kirjoitusvirheiksi ja ehkä myös vartalon hahmottamisen virheiksi (21) ja sekä vokaaliharmoniaongelmiksi (22); jälkimmäinen ongelma on tavallinen vironkielisilläkin oppijoilla. PITÄÄ-verbin osalta keskityimme konsonanttivaihtelussa ja sitä läheisesti muistuttavassa passiiviin tunnuksen konsonanttien vaihtelussa esiintyviin virheisiin, koska ne aiheuttavat runsaasti ongelmia indoeurooppalaisille kielenoppijoille ja käänteinen astevaihtelu myös sukulaiskielen oppijoille (ks. Kaivapalu 2005: 51). Astevaihteluvirheitä voi esiintyä sekä verbin vartalossa että passiivimuodon tunnuksesta, mutta koska passiivimuodot olivat molemmissa aineistoissa hyvin harvinaisia, löytyi pääosa virheistä vartalosta (23, 24).¹⁰

- (21) Joka osastossa on kaksi huonetta kahden **ihemesen** varten, keittiö ja kylpy. (VE0175a)
- (22) Minun perhessa on kolme **ihmista**- minä, minun äiti ja minun veli. (VI0277)
- (23) Hän **pidää** musiikista, mutta hän ei laula. (VI0116A)
- (24) Minä **pitän** opiskelemisesta Potznanissa. (PU0036)

Toisin kuin useissa aikaisemmissa tutkimuksissa (esim. Kaivapalu 2005; Kaivapalu & Eslon 2011) ei meidän analyysimme mukaan sukulaiskielellä ole positiivista vaikutusta tuotoksen laatuun. Tapauksessa IHMINEN ICLFI-multi-aineistossa virheellisiä muotoja on 6,1 prosenttia, kun taas ICLFI-virossa niitä on 5,9 prosenttia. Myös tapauksessa PITÄÄ prosenttiosuudet ovat liki samat: ICLFI-multi-aineistossa 6,6 % ja ICLFI-virossa 6,4 %. Käyttämiemme aineistojen ja tutkimamme kahden tapauksen perusteella emme siis voi väittää, että äidinkielenään sukukieltä puhuvat kielenoppijat hallitsevat taivutusmuodot muita kielen-

¹⁰ Oletuksemme on, että vaikka virolaisilla kielenoppijoilla on oman vastaavan järjestelmänsä johdosta jonkinlainen käsitys suomen konsonanttivaihtelusta, voivat vaihtelun eroavaisuudet aiheuttaa heille ongelmia. On myös muistettava, että aina ei voi tehdä eroa taivutusvirheiden ja pelkkien kirjoitusvirheiden välille.

oppijoita paremmin. On toki otettava huomioon, että aineistot eivät ole täysin vertailukelpoisia taitotasoiltaan: ICLFI-viro-aineistossa on erittäin paljon tasoilla B1-olevia, meidän luokituksessamme alkeistasolle sijoittuvia kielenoppijoita, mikä näkyy erityisesti PITÄÄ-verbin tuotoksissa.

8. Tuloksia ja pohdintaa

Tutkimustulostemme perusteella voimme väittää, että fraseologia ja morfologinen priming ovat yhteydessä toisiinsa. Koska kielenoppijat tuottavat epätyypillisiä fraseologisia yksiköitä, myös heidän paradigmaattinen morfologinen priminginsa eroaa natiivipuhujien primingista.

Tutkimuksemme osoittavat, että suomea vieraana kielenä opiskelevat kielenoppijat eivät hallitse fraseologisen yksikön ytimen paradigmaattista morfologista primingia samalla tavalla kuin natiivipuhujat. Tämä näkyy erityisesti siinä, että kielenoppijoiden tuottamien erilaisten taivutusmuotojen frekvenssit eroavat selvästi natiivipuhujien tuottamista muodoista. Sekä tapauksessa IHMINEN että PITÄÄ löytyy tilastollisesti merkitseviä yli- ja aliedustumisia. Oppijankielen fraseologisen yksikön ytimen paradigmaattinen valinta siis eroaa kohdekielestä.

Myös ytimen ja sen kerasanojen muodostaman kokonaisuuden – eli fraseologisen yksikön – morfologinen priming on erilaista oppijankielessä kuin kohdekielessä.

Yhtenä mahdollisena syynä epätyypilliseen morfologiseen primingiin on kielenoppijoiden pyrkimys välttää tiettyjä muotoja. Tämä näkyy esimerkiksi nominin *ihminen* suosimisessa morfofonologisesti monimutkaisen passiivimuodon sijasta. Kohdekielestä poikkeavat epätyypilliset primingit voivat olla myös rakenteellisesti hyvinmuotoisia ilmauksia kuten *perheessäni on # ihmistä*, jotka natiivikielessä kuitenkin ovat hyvin harvinaisia. Kielenoppijoiden käyttämien PITÄÄ-verbin eri merkitysten variaatio on selvästi kapeampi kuin natiivipuhujien ja tietyt, usein konkreettiset ja opetuksessa painotetut rakenteet ja sitä myötä merkitykset korostuvat muiden kustannuksella: tutkimuksemme kielenoppijoiden

tuottamat *pidän*-muodot liittyvät selvästi yleisimmin elatiivitäydennyksiseen merkitykseen ‘tykätä’. Mielenkiintoista kuitenkin on, että verbin muoto vaikuttaa merkitysvariaatioiden käyttöön: muodolla *pitää* on positiivinen priming verbin fraasikäyttöön kun taas muodolla *pidän* priming on negatiivinen. Näin pilottitutkimuksemme vahvistaa Sinclairin (1991: 8) ja Stubbsin (2001: 17) käsitystä, että leksikaalisia (fraseologisia) yksiköitä ei pitäisi tarkastella yksinomaan lemman tasolla, elleivät lemman eri taivutusmuotojen syntagmaattiset valinnat ole yhteneviä. Erityisen tärkeää tämä on silloin, kun tutkimuskohteena on kieli, jolla on rikas morfosyntaktinen taivutusjärjestelmä.

Aikaisemmista tutkimuksista poiketen analyysimme osoittaa, että kielenoppijan äidinkielellä ei näytä olevan suoranaista vaikutusta käytettyjen muotojen virheettömyyteen – tosin tutkimamme kahden tapauksen tulokset ovat kiinnostavalla tavalla vastakkaisia: tapauksessa IHMINEN sukukielen puhujien priming on lähempänä natiivipuhujien kielenkäyttöä kuin äidinkielenään ei-sukukieltä puhuvien kielenoppijoiden, mutta tapauksessa PITÄÄ tulos on päinvastainen. On kuitenkin otettava huomioon, että tulosten taustalla voi olla monia muitakin muuttujia. Käytetyt oppikirjat, opetusmenetelmät, opettajan äidinkieli sekä tietysti tehtävänannot ja niiden ohjeistus voivat vaikuttaa sekä fraseologisten yksiköiden muodostumiseen että tuotettuihin morfologisiin muotoihin ja sitä myötä kielenoppijoiden tuottamaan morfologiseen primingiin. Yhtä kaikki tämä pilottiluonteinen kahden tapauksen tarkastelu paljastaa fraseologisuuden ja morfologisen primingin keskinäisen vaikutuksen olomassaolon ja osoittaa, että morfologinen priming vaatii runsaasti lisätutkimusta siinä missä ylipäättään fraseologinen osaaminen ja siinä kehittyminenkin.

Kiitokset

Tutkimusta ovat tukeneet Riksbankens Jubileumsfond (F10-0421:1) ja Oulun yliopisto. Kiitämme myös nimettömiä arvioijia hyödyllisistä kommentteista.

Lyhenteet

1	ensimmäinen persoona
3	kolmas persoona
GEN	genetiivi
INF	infinitiivi
NOM	nominatiivi
NP	nominilauseke
PL	monikko
PTV	partitiivi
SG	yksikkö

Lähteet

- Aston, Guy, Lou Burnard 1998. *The BNC Handbook. Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.
- Dose, Stephanie 2010. *Patterns of growing standardisation and interference in interpreted German discourse*. MA Dissertation. University of South Africa. <http://uir.unisa.ac.za/xmlui/handle/10500/4710?show=full> (17.3.2012).
- Ellis, Nick C. 2008. *Phraseology: The periphery and the heart of language*. – Fanny Meunier, Sylviane Granger (Eds.). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins, 1–13.
- Granger, Sylviane 1996. *From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora*. – Karin Aijmer, Bengt Altenberg, Mats Johansson (Eds.). *Languages in Contrast. Text-based Cross-linguistic Studies*. Lund: Lund University Press, 37–51.
- Granger, Sylviane 1998. *Prefabricated patterns in advanced EFL writing: Collocations and formulae*. – Anthony Paul Cowie (Ed.). *Phraseology: Theory, Analysis and Applications*. Oxford: Oxford University Press, 145–160.
- Hasselgren, Angela 1994. *Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary*. – *International Journal of Applied Linguistics* 4 (2), 237–260. <http://dx.doi.org/10.1111/j.1473-4192.1994.tb00065.x>
- Hoey, Michael 2005. *Lexical Priming. A New Theory of Words and Language*. London: Routledge.
- Hoey, Michael 2007. *Lexical priming and literary creativity*. – Michael Hoey, Michaela Mahlberg, Michael Stubbs, Wolfgang Teubert (Eds.). *Text, Discourse and Corpora. Theory and Analysis*. London: Continuum, 7–29.

- ISK = Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja-Riitta Heinonen, Irja Alho 2004. Iso suomen kielioppi. Suomalaisen Kirjallisuuden Seuran toimituksia 950. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Jantunen, Jarmo H. 2001. *Tärkeä seikka ja keskeinen kysymys*: Mitä korpuslingvistinen analyysi paljastaa lähisyronyymeista? – *Virittäjä* 105 (2), 170–192.
- Jantunen, Jarmo H. 2004a. Synonymia ja käännössuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin. Joensuun yliopiston humanistisia julkaisuja 35. Joensuu: Joensuun yliopisto.
- Jantunen, Jarmo H. 2004b. Unypical patterns in translations. *Issues on corpus methodology and synonymity*. – Anna Mauranen, Pekka Kujamäki (Eds.). *Translation Universals. Do they Exist?* Amsterdam: John Benjamins, 101–126.
- Jantunen, Jarmo H. 2007. Oppijansuomen piirteitä korpusvetoisesti. – Pirkko Muikku-Werner, Ossi Kokko, Hannu Remes (Toim.). *Virsu* 3. Suomalais-ugrilaisia kohdekieliä ja kontakteja. *Studies in Languages* 42. Joensuu: Joensuun yliopisto, 69–83.
- Jantunen, Jarmo H. 2009a. Ei pelkästään mielikuvituksen puutteen vuoksi. Kieliaineistojen systemaattinen käyttö kielentutkimuksessa. – *Virittäjä* 113 (1), 101–113.
- Jantunen, Jarmo H. 2009b. *Minulla on aivan paljon rahaa*. Fraseologiset yksiköt suomen kielen opetuksessa. – *Virittäjä* 113 (3), 356–381.
- Jantunen, Jarmo H. 2011a. Avainsana-analyysi annotoiduin oppijankieliaineiston tutkimisessa: alustavia havaintoja. – Esa Lehtinen, Sirkku Aaltonen, Merja Koskela, Elina Nevasaari, Mariann Skog-Södersved (Toim.). *AFinLA-e: Soveltavan kielitieteen tutkimuksia* 3, 48–61. <http://ojs.tsv.fi/index.php/afinla> (16.3.2012).
- Jantunen, Jarmo H. 2011b. Kansainvälinen oppijansuomen korpus (ICLFI): typologia, taustamuuttajat ja annotointi. – *Lähivördlusi. Lähivertailuja* 21, 86–105. <http://dx.doi.org/10.5128/LV21.04>
- Jantunen, Jarmo H. 2012. Korpusvetoinen tekstilajianalyysi: sanalistat ja genreavainsanat. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä, Mikko Lounela (Toim.). *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus.
- Kaivapalu, Annekatrin 2005. Lähdekieli kielenoppimisen apuna. *Jyväskylä Studies in Humanities* 44. Jyväskylä: Jyväskylän yliopisto.
- Kaivapalu, Annekatrin, Pille Eslon 2011. Onko lähisukukielen vaikutus suomen ja viron omaksumiseen symmetristä. Korpuspohjaisen tutkimuksen tuloksia

- ja haasteita. – Lähivertailuja. Lähivördlusi 21, 132–151. <http://dx.doi.org/10.5128/LV21.06>
- Kallioranta, Otto 2009. Paljon-adverbin kollokointi oppijansuomessa: Korpusvetoinen tutkimus. Pro gradu -tutkielma. Oulun yliopisto, suomen kieli.
- Kaszubski, Przemek 2000. Selected Aspects of Lexicon, Phraseology and Style in the Writing of Polish Advanced Learners of English: A Contrastive, Corpus-Based Approach. Poznań: Adam Mickiewicz University.
- KS = Kielitoimiston sanakirja 2008. Kotimaisten kielten tutkimuskeskuksen julkaisuja 149. Helsinki: Kotimaisten kielten tutkimuskeskus ja Kielikone.
- Kubáčková, Jana 2009. Keeping Czech in check: A corpus-based study of generalization in translation. – SKASE Journal of Translation and Interpretation 4 (1), 33–51. http://www.skase.sk/Volumes/JTI04/pdf_doc/03.pdf (17.3.2012).
- Martin, Maisa, Sanna Mustonen, Nina Reiman, Marja Seilonen 2010. On becoming an independent user. – Inge Bartning, Maisa Martin, Ineke Vedder (Eds.). Communicative Proficiency and Linguistic Development: Intersections Between SLA and Language Testing Research. Eurosla Monographs Series 1. Eurosla, 57–79.
- Mauranen, Anna 2000. Strange strings in translated language. A study on corpora. – Maeve Olohan (Ed.). Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects. Manchester: St. Jerome, 119–141.
- McNamara, Timothy P. 2005. Semantic Priming. Perspectives from Memory and Word Recognition. New York: Psychology Press. <http://dx.doi.org/10.4324/9780203338001>
- Meunier, Fanny, Sylviane Granger 2008. Phraseology in Foreign Language Learning and Teaching. Amsterdam: John Benjamins.
- Neely, James H. 1991. Semantic priming effects in visual word recognition. A selective review of current findings and theories. – Derek Besner, Glyn W. Humphreys (Eds.). Basic Processes in Reading: Visual Word Recognition. Hillsdale, NJ: Erlbaum, 264–336.
- Nesselhauf, Nadja 2003. The use of collocations by advanced learners of English and some implications for teaching. – Applied Linguistics 24 (2), 223–242. <http://dx.doi.org/10.1093/applin/24.2.223>
- Nesselhauf, Nadja 2005. Collocations in a Learner Corpus. Amsterdam: John Benjamins.
- Nesselhauf, Nadja 2009. Co-selection phenomena across New Englishes. Parallels (and differences) to foreign learner varieties. – English World-Wide 30 (1), 1–25. <http://dx.doi.org/10.1075/eww.30.1.02nes>

- Scott, Mike 2004. WordSmith Tools version 4. Oxford: Oxford University Press.
- Scott, Mike, Christopher Tribble 2006. Textual Patterns. Key Words and Corpus Analysis in Language Education. Studies in Corpus Linguistics 22. Amsterdam: John Benjamins.
- Seilonen, Marja 2011. Olemattoman ilmaiseminen. – Esitelmä Suomen kielen ja kulttuurin opettajien opinto- ja neuvottelupäivillä Kuopiossa 4.8.2011.
- Sinclair, John 1991. Corpus, Concordance, Collocation. Oxford: Oxford University Press.
- Sinclair, John 1996. The Search for Units of Meaning. – Textus IX, 75–106.
- Stubbs, Michael 2001. Words and Phrases. Corpus Studies of Lexical Semantics. Oxford: Blackwell.
- Tognini-Bonelli, Elena 1996. Corpus Theory and Practice. Birmingham: TWC.
- Tsui, Amy B. M. 2004. What teachers have always wanted to know – and how corpora can help. – John Sinclair (Ed.). How to Use Corpora in Language Teaching. Studies in Corpus Linguistics 12. Amsterdam: John Benjamins, 39–61.
- Tsui, Amy B. M. 2005. ESL teachers' questions and corpus evidence. – International Journal of Corpus Linguistics 10 (3), 335–356. <http://dx.doi.org/10.1075/ijcl.10.3.03tsu>
- Wray, Alison 1999. Formulaic language in learners and native speakers. – Language Teaching 32 (4), 213–231. <http://dx.doi.org/10.1017/S0261444800014154>
- Wray, Alison 2000. Formulaic sequences in second language teaching. Principle and practice. – Applied Linguistics 21 (4), 463–489. <http://dx.doi.org/10.1093/applin/21.4.463>

Jarmo Harri Jantunen

Oulun yliopisto, suomi toisena ja vieraana kielenä
PL 1000
90014 Oulun yliopisto, Finland
jarmo.jantunen@oulu.fi

Sisko Brunni

Oulun yliopisto, suomen kieli
PL 1000
90014 Oulun yliopisto, Finland
sisko.brunni@oulu.fi

Morphological priming and phraseology in second language acquisition: A corpus-study on learner Finnish

JARMO HARRI JANTUNEN

University of Oulu

SISKO BRUNNI

University of Oulu

This article discusses morphological priming in the second language acquisition. Morphological priming is a characteristic of both the core and cotextual items in a phraseological unit. It occurs when a word is frequently encountered in certain morphophonological forms. Similarly to lexical priming on the whole (e.g. collocations and other cotextual qualities), it poses challenges for language learners. The paper focuses on atypicalities in morphophonological forms and, in addition, describes errors in inflection. It is hypothesized that learners of Finnish have problems in morphological priming, and that learners, whose mother tongue is closely related to the target language, in which there exists inflection, produce more target like phraseological units.

Keywords: morphological priming; lexical priming; phraseology; learner Finnish; crosslinguistic transfer; corpus analysis