

# Oppikirjan sanaston vaikutuksesta ruotsinkielisten alkeistason suomenoppijoiden kirjallisiin tuotoksiin

TUIJA MÄÄTTÄ

Uumajan yliopisto

**Tiivistelmä.** Artikkelissa pohditaan oppikirjan sanaston mahdollista vaikutusta ruotsinkielisten alkeistason suomenoppijoiden kirjallisiin tuotoksiin. Tutkimus on korpuspohjainen ja sisältää kaksi korpusta: suomenoppijoiden kirjallisten tuotosten korpuksen ja oppikirjan viiden kappaleen sanastojen korpuksen. Suomenoppijoiden korpus kuuluu osana Kansainväliseen oppijansuomen korpuksen. Molempien korpusten sanat ovat siinä muodossa kuin ne ovat esiintyneet alkuperäisissä lähteissä. Tutkimusmetodi on komparatiivinen ja kvantitatiivinen. Vertaamalla yhtäältä oppijoiden käyttämiä, toisaalta oppikirjassa esiintyviä sanoja ja sanamuotoja toisiinsa tarkastellaan, missä määrin oppijat ovat hyödyntäneet oppikirjaa. Mikäli samoja sanoja ja muotoja on löydettävissä, kyseessä voi olla oppikirjan vaikutus oppijoiden kieleen. Artikkelissa tarkastellaan myös eri sanaluokkien esiintymistä tuotoksissa ja oppikirjassa. Lisäksi kartoitetaan oppijoiden tuotoksissaan käyttämien sanojen yleisyyttä ja esiintymistiheyttä.

**Avainsanat:** korpuspohjainen tutkimus; sanasto; oppijansuomi; sanaluokat; sanafrekvenssi; suomi; ruotsi

## 1. Johdanto

Artikkelissani esittelen tutkimustani, jossa pyrin selvittämään leksikon tasolla, miten oppikirjan sanasto mahdollisesti vaikuttaa ruotsinkielisten suomenoppijoiden kirjoittamiin teksteihin. Tarkasteltavana on suomen kielen alkeiskurssilla Uumajassa käyttämämme kirja “Suomen mestari 1” (Gehring & Heinzmann 2010). Uumajan yliopiston alkeiskurssi “Finska, nybörjarkurs steg 1–2” pidetään syyslukukaudella, ja sen laajuus on 15 opintopistettä. Kontaktiopetusta kurssilla on 5–6 tuntia viikossa, opetuskieli on ruotsi. Oppikirjan lisäksi käytössä on runsaasti opettajien tuottamaa materiaalia, ja suullisen viestinnän osiossa hyödynnetään myös äänitteitä ja verkkoa. Kurssi sisältää neljä osiota: suomen kielen rakenne, kirjallinen viestintä, suullinen viestintä ja sanasto. Sanaston osio on täysin verkkopohjainen.

Kurssin aikana opiskelijoiden sanasto karttuu kaikkien neljän kurssiosion kautta. Oppikirja ei siis ole oppijoiden ainoa lähde, josta he ammentavat sanoja. Muita sanaston lähteitä ovat esimerkiksi Puron (1999: 9) mukaan oppikirjojen ulkopuolinen aineisto, sanakirjat ja ympäröivä kieliyhteisö, joka luonnollisesti puuttuu opiskeltaessa suomea maan rajojen ulkopuolella. Monelle oppijalle oppikirja on ainoa väylä suomen kieleen ja suomalaiseen kulttuuriin. Näin ollen oppikirja saa jonkinlaisen auktoriteettiaseman (Tanner 2008: 167).

Nykyopiskelijat ovat ahkeria verkon käyttäjiä. Esimerkiksi suomenkielisiä verkkokursseja, harjoitusmateriaaleja, sanastoja ja sanakirjoja on runsaasti tarjolla. Hyvin motivoituneet opiskelijat hyödyntävät näitäkin lähteitä. Unohtaa ei sovi myöskään sitä, että kaikki oppijat eivät ole ihan todellisia vasta-alkajia tullessaan alkeiskurssille vaan heillä voi olla erilaista tietoa ja eriasteisia taitoja suomen kielestä muuta kautta.

Suomi toisena ja vieraana kielenä -opetuksessa on melko pitkän ajan painotettu kieliopin ja sääntöjen opettamista ja oppimista (ks. esim. Järventausta 2004: 25). Kielen kommunikatiivinen puoli on jäänyt taka-alalle. Muun muassa Milton (2009: 1) ja Meara (2009: ix) ovat todenneet, että toisen kielen omaksumisessa sanastoon on kiinnitetty

paljon vähemmän huomiota kuin muihin kielen alueisiin. On itsestään selvää, että ilman sanoja ja sanatuntemusta ei ole minkäänlaisia kommunikaatiomahdollisuuksia. Järventaustan (2004: 26) mukaan sanastoa opetettaessa ja opittaessa pitäisi olla tavoitteena natiivin kielenkäyttäjän kompetenssia vastaava leksikaalinen tieto.

Kielenoppijoiden kirjoituksissaan käyttämään sanastoon tulisi kiinnittää enemmän huomiota. Monet opettajat tarkastelevat enemmän tekstien kieliopillista korrektiutta kuin sanastoa (Schmitt 2000: 155). Tutkimuksissa on jo alettu suunnata huomio sanaston opettamiseen ja oppimiseen (suomen osalta ks. esim. Masonen 2003). Tutkimuksissa pohditaan, millaista sanastoa pitäisi opettaa ja millaisia sanoja oppijat oppivat ja omaksuvat. Lundberg (2011: 68) pitää hyvin tärkeänä, että yksittäisiä sanoja ei eristetä vaan että ne tulisi asettaa sellaisiin yhteyksiin, jotka ovat oppijoille mielenkiintoisia ja relevantteja. Zimmerman ja Schmitt (2005: 2) korostavat sanojen frekvenssiä ja toteavat, että frekventeimmät sanat ovat käyttökelpoisimpia: mitä suurtaajuisempi sana on, sitä hyödyllisempää sen opettaminen on. Frekventit sanat ovat frekventtejä yksinkertaisesti siksi, että ne esiintyvät paljon kielessä ja hyvin vaihtelevissa tilanteissa. Siksi sanat ovat arvokkaita mitä erilaisimmissa konteksteissa.

Schmitt ja Carter (2000: 6) on leksikaalisten fraasien puolestapuhuja. Hänen näkemyksensä mukaan leksikaaliset fraasit ovat sujuvan kielen tuottamisen avainelementtejä, ja niillä on osansa sanaston oppimisessa. Myöhemmässä tutkimuksessaan Schmitt käyttää leksikaalisista fraaseista termiä *formulaic sequences*; kutsun niitä tässä artikkelissa sanasekvensseiksi. Conklin ja Schmitt (2008: 73) toteavat sanasekvenssien olevan enemmän kuin sanojen ketjuja, jotka on kytketty toisiinsa kollokationaalisin keinoin. Tutkimuksessaan sanasekvenssien lukemisesta he havaitsivat, että ne nopeuttivat sekä natiivien että L2-oppijoiden lukemista (Conklin & Schmitt 2008: 84).

Paljon kielen kommunikatiivisesta sisällöstä on sidoksissa fraseologiin ilmauksiin. Sanasekvenssejä on paljon, ja funktionaalisesti ne ovat tärkeitä kirjoitetuissa teksteissä. Tästä huolimatta L2-oppijoille niiden käyttö on ongelmallista. Tämä ilmenee muun muassa siten, että

rajoitettua määrää hyvin tunnettuja sanasekvenssejä käytetään erittäin paljon. (Li & Schmitt 2009: 85.)

Myös Aronen ja Kuusk (2010: 107) ovat pohtineet sanaston opettamista. Heidän mielestään on yleinen ajattelutapa, että oppijat itse kartuttavat sanavarastoaan. Usein oppijat saavat luettavakseen sanalistoja, ja oletetaan, että he näin siirtävät sanoja aktiivisesti hallittavaan sanavarastoon. Aronen ja Kuuskin mukaan sanaston opetuksessa on tärkeää kielitieto ja oppimistapojen vaihtelu. He ovat tutkineet suomen ja viron sanastoa, sen opettamista ja erityishuomiota tarvitsevia seikkoja, kuten kollokaatioita, näiden kahden kielen erimerkityksisiä sanoja sekä monimerkityksisiä sanoja (2010: 109). Mielestäni nämä seikat eivät ole ainoastaan kahden lähikielen opetuksen ongelmakohtia vaan niihin pitäisi panostaa kaikessa S2-opetuksessa, olkoon suomenoppijan lähdekieli mikä hyvänsä.

Kirjoittaminen on yksilöllistä. Jos kirjoittaminen ei omalla äidinkielelläkään suju, on todennäköistä, että myös toisella kielellä kirjoittaminen on haastava tehtävä. Grönholmin (1993: 411) mukaan kirjoittaminen on kognitiivinen prosessi, joka vaatii oppijalta paljon. Tottumattoman kirjoittajan pitää kiinnittää huomiota itse kirjoittamiseen, ja kirjoittaessaan toisella kielellä hänen täytyy paneutua oikeinkirjoitukseen. Tutkimuksissa on tullut esiin (ks. esim. Krapels 1990: 37–41), että puutteet, joita saattaa esiintyä toisella kielellä kirjoittavien tuotoksissa, eivät johdu kielellisen kompetenssin puutteellisuudesta. Kyse on paremminkin siitä, että itse kirjoitustaitoa ei hallita tarpeeksi hyvin.

Krapels on havainnut, että äidinkielellä ja toisella kielellä kirjoittavien välillä on paljon yhteistä. Kirjoittamisprosesseissa on kuitenkin eroavaisuuksia. Krapelsin mukaan L2-kirjoittajat kirjoittavat enemmän kuin L1-kirjoittajat. L2-kirjoittajat eivät kiinnitä niin paljon huomiota virheisiin kuin L1-kirjoittajat, ja L1-kirjoittajat korjaavat tuotoksiaan mutta L2-kirjoittajat eivät (1990: 44–45). Uudemmissa tutkimuksissa (Schoonen ym. 2009: 86–88) on saatu päinvastaisia tuloksia. Niiden mukaan L2-kirjoittajat tarkistavat enemmän tekstiensä kieltä ja myös typografiaa kuin L1-kirjoittajat.

Käsillä olevan korpuspohjaisen tutkimuksen kautta olen etsinyt vastauksia seuraaviin tutkimuskysymyksiin:

- Onko suomenoppijoiden kirjoittamien tekstien sanastossa merkkejä oppikirjan sanaston vaikutuksesta?
- Minkä tyyppistä vaikutusta on nähtävissä?
- Millainen sanaluokkien jakauma on suomenoppijoiden tuotoksissa?
- Mitkä sanat ovat frekventeimmät?

## 2. Aineisto ja tutkimusmenetelmä

Tutkimus perustuu kahteen digitaaliseen aineistoon: suomenoppijoiden tuotosten korpukseen ja opetusharjoittelijan laatimaan oppikirjan kappaleiden sanastojen korpukseen. Syyslukukaudella 2010 17 oppijaa kirjoitti kukin kolme tekstiä. Ne kirjoitettiin itsenäisesti kotona, ja erilaisten apuvälineiden käyttö oli sallittua. Tehtävänanto oli seuraavan kaltainen:

Teksti A: *Kirjoita sähköpostiviesti ystävällesi.*

Teksti B: *Kuvaile fantasiahenkilö tai oikea henkilö.*

Teksti C: *Kerro tavallisesta päivästäsi elämässäsi.*

Tekstien otsikot ja aihepiirit seuraavat oppikirjan lukutekstejä. Teksti A on kirjoitettu oppikirjan kolmannen kappaleen jälkeen. Oppikirjan lukutekstinä on juuri sähköpostiviesti, jossa viestin kirjoittajana on Suomessa opiskeleva brasilialainen opiskelija. Viestin vastaanottaja on Brasiliassa asuva suomalainen mies.

Oppikirjan neljännessä kappaleessa opetellaan adjektiiveja ja niiden käyttöä. Kappaleessa tutustutaan myös omistusrakenteeseen. Lukutekstissä eri henkilöt kertovat lyhyesti itsestään ja perheestään. Tekstissä B on tarkoituksena harjoitella adjektiiveja ja omistusrakennetta henkilökuvaailun kautta.

Teksti C on tekstityypiltään kertomus. Oppijan tehtävänä on kertoa ja kuvailla, mitä hän tekee päivänsä aikana ja mitä milloinkin tapahtuu.

Oppikirjan viidennen kappaleen lukutekstissä kerrotaan erään perheen arkipäivästä aamusta iltaan.

Kirjallisten tuotosten korpus sisältää 6 250 sanetta. Tekstien kesken saneet jakautuvat seuraavasti: teksti A: 1 722 sanetta, teksti B: 2 599 sanetta ja teksti C: 1 929 sanetta. Aineisto kuuluu osana ruotsinkielisten suomenoppijoiden osakorpukseseen, joka puolestaan on osakorpuksena Kansainvälisessä oppijansuomen korpuksessa, ICLFI (Jantunen 2011: 86–105; Määttä 2010a: 16–17; Määttä 2010b: 18–20; Jantunen & Pilttonen 2009: 449–458). ICLFI-korpuksen tekstit on määritelty tekstityypin mukaan.

“Suomen mestari 1” on yksikielinen, suomenkielinen, oppikirja, joka kirjan tekijöiden mukaan on “aikuisille tarkoitettu oppikirja suomen kielen opiskeluun alkeistasolla” (Gehring & Heinzmann 2010: 6). Syksyllä 2010 Uumajassa toiminut Cimon harjoittelija Minna Mattanen laati oppikirjan kappaleisiin suomi–ruotsi-sanastot ja fraasisanastot. Sanastot sisältävät kaikki sanat, jotka esiintyvät lukuteksteissä, dialogeissa ja erilaisissa harjoituksissa, ja sanat ovat juuri siinä muodossa kuin ne esiintyvät oppikirjassa. Ruotsinnoksen lisäksi sanastoihin on merkitty sanojen sanaluokat. Sanastot ovat kurssin verkkosivuilla.

Opiskelijat toimittivat tuotoksensa word-tiedostoina sähköpostiliitteenä opettajalle. Jotta tekstejä voi tutkia koneellisesti käyttämällä *WordSmith Tools* -tietokoneohjelmaa, tekstit on muunnettu txt-tiedostoiksi samoin kuin sanastotkin, jotka olivat alun perin xls-tiedostoja. Molemmista korpuksista on sen jälkeen otettu aakkoselliset sanalistat *WordList*-työkalulla. Sanojen frekvenssitutkimusta varten oppijoiden tuotoksista on otettu erikseen frekvenssisanalistat.

Sanaluokkien tutkiminen ei ole ollut mahdollista koneellisesti, koska ICLFI-korpus on annotoitu kieliopillisesti vasta osittain (Jantunen 2011: 89).

Tutkimuksessa käytetään kvantitatiivista menetelmää. Molemmista korpuksista saatuja tuloksia myös vertaillaan keskenään. Koska molempien korpusten aakkosellisissa sanalistoissa sanat ovat siinä muodossa kuin ne ovat esiintyneet lähdeaineistoissa, listojen vertailu on mahdol-

lista. Vertailutyö on tehty manuaalisesti. Tällä tavalla saadaan selville, mitkä sanat ja sanamuodot ovat tuotoksissa juuri samat kuin oppikirjan sanastoissa.

Sanojen frekvensseistä saadut tiedot on koostettu suoraan frekvenssisanalistoista. Koska oppijoiden tuotosten osakorpus on vielä kieliopillisesti annotoimaton, sanaluokat on merkitty tuotosten sanalistoihin käsin.

Tutkimusmenetelmän löytäminen oli vaikeaa, ja suomi vieraana kielellä -tutkimuskentältä en ole löytänyt vastaavaa tutkimusta kuin käsillä oleva. Sundman (2011: 327–330) on tutkinut suomenkielisten 8. ja 9. vuosikurssin oppilaiden ruotsinkielisiä kirjallisia tuotoksia käyttäen vertailevaa tutkimusmetodia. Tutkimuskohteena ovat sanasekvenssit (engl. *formulaic sequences*). Sundman paitsi vertailee sekvenssien esiintymiä eri oppilaiden tuotoksissa myös tarkastelee niiden sijoittumista yhtäältä oppilaiden teksteihin, toisaalta oppimateriaaliin. Sundmanin tarkoituksena on identifioida sellaisia ilmauksia, joita oppilaat ovat ottaneet tuotoksiinsa melkein muuttumattomissa muodoissa oppimateriaalista. Oppikirjojen kielestä sinänsä, esimerkiksi niiden dialogeista (Karppinen 1996; Tanner 2012), tutkimusta on tehty, ja oppikirjoja on vertailtu keskenään. Sitä vastoin oppikirjan ja opiskelijan tuotosten väliseen vertailuun ei ole kiinnitetty huomiota.

Schmittin (2010: 12–13) mukaan sanaston tutkimuksessa korpusanalyysi on tärkeä tutkimustyökalu ja se on merkittävästi kehittänyt tätä tieteenalaa. Korpuksia käytetään paljon erilaisissa frekvenssitutkimuksissa, ja Schmittin mielestä yksi tärkeimmistä sanaston piirteistä on juuri frekvenssi, joka vaikuttaa useimpiin tai kaikkiin aspekteihin leksikaaliosassa prosessissa ja omaksumisessa.



### 3. Kirjallisten tuotosten ja oppikirjan sanastojen vertailua

Tutkimuksen kohteena olevat kolme tekstiä suomenoppijat kirjoittivat tutustuttuaan oppikirjan kolmanteen, neljänteen ja viidenteen kappaleeseen. Jotta tuotoksissa esiintyvistä oppikirjan sanastosta käytettyjen sanojen esiintymisistä saa tarkempia tuloksia, on tutkittava myös, mitä sanoja kussakin tuotoksessa on tehtävää edeltävistä oppikirjan kappaleista mukana (teksti A kappaleet 1–3, teksti B kappaleet 1–4, teksti C kappaleet 1–5). Taulukossa 1 esitän esiintyneiden sanojen lukumäärät ja prosenttiosuudet.

**TAULUKKO 1.** *Oppijoiden tuotoksissa käytettyjen oppikirjan sanojen määrät*

| Kirjan kappale | Teksti A  |             | Teksti B  |             | Teksti C   |             | Summa |      |
|----------------|-----------|-------------|-----------|-------------|------------|-------------|-------|------|
|                | f         | %           | f         | %           | f          | %           | f     | %    |
| 1              | 81        | 11,5        | 30        | 4,6         | 51         | 5,9         | 162   | 21,4 |
| 2              | 75        | 10,6        | 52        | 8,0         | 58         | 6,7         | 185   | 24,4 |
| 3              | <b>89</b> | <b>12,6</b> | 66        | 10,2        | 51         | 5,9         | 206   | 27,3 |
| 4              | –         | –           | <b>70</b> | <b>10,8</b> | 26         | 3,0         | 96    | 12,7 |
| 5              | –         | –           | –         | –           | <b>109</b> | <b>12,5</b> | 109   | 14,4 |
| Summa          | 245       | 34,7        | 218       | 33,6        | 295        | 34,0        | 758   | 100  |

Taulukko 1 osoittaa, että kaikissa kolmessa tekstissä oppikirjan sanoja on eniten juuri siitä kappaleesta, johon tutustumisen jälkeen tekstit on tuotettu. Tulos ei hämmästyttä, kun ajattelee kirjoitustehtävien tarkoitusta. Koska uutta sanastoa on juuri opittu, sitä on luonnollisesti käytetty hyväksi kirjoitustehtävissä, jotka aiheiltaan muistuttavat kirjan kappaleiden tekstejä.

Verrattaessa oppikirjasta otettujen sanojen prosentuaalista määrää eri teksteissä huomataan, että se on hyvin tasainen. Keskimäärin 34 % tuotoksien sanoista voi sanoa olevan peräisin oppikirjasta. Määrän olisi



voinut olettaa olevan korkeampi, koska oppikirjan teksteihin tutustutaan opetuksessa tarkasti. Myös opettajien laatimissa suullisissa ja kirjallisissa harjoituksissa kirjan sanasto toistuu jatkuvasti.

Taulukko 1 antaa tuotoksien sanastosta sen kuvan, että oppikirjan kolmas kappale sisältää ikään kuin jonkinlaista ydinsanastoa, jota oppijat tarvitsevat useissa yhteyksissä: kappale sisältää adjektiivien lisäksi erilaisia sää- ja ajanilmauksia, ja verbimäärä lisääntyy astevaihtelun alaisilla verbeillä. Masonen (2003: 94) on todennut, että oppijoille pitäisi antaa sellaista ainesta, jonka he tuntevat tärkeäksi eri oppimisvaiheissa. Tällä tavalla synnytetään osaamisen tunnetta ja oppijat motivoituvat. Moni suomenoppija aloittaa suomen opinnot pystyäkseen kommunikoidaan suomeksi. Siksi arkielämään liittyvä sanasto onkin keskiössä opintojen alussa, jolloin tarvitaan konkreettisia substantiiveja ja verbejä.

Tekstin A kohdalla merkittävää on se, että myös ensimmäisestä ja toisesta kappaleesta on mukana sanoja melkein pä yhä paljon kuin kolmannestakin. Syy lienee se, että teksti A on oppijoiden ensimmäinen itsenäinen kirjoitustehtävä, ja se on tuotettu vain muutaman opintoviikon jälkeen.

Seuraavassa on esimerkkejä siitä, miten suomenoppijat (SO) ovat käyttäneet tekstissä A yksittäisiä sanoja, sanontoja ja kokonaisia lauseitakin, jotka ovat oppikirjan (OK) kolmessa ensimmäisessä kappaleessa.

- (1) OK: Olen suomen kurssilla maanantaisin, keskiviikkoisin ja torstaisin.
- (2) SO: Olen suomen kurssilla joka viikko Maanantaisin ja Torstaisin.
- (3) OK: Aurinko paistaa.
- (4) SO: Aurinko paistaa.
- (5) OK: Tänään minä en kirjoita sinulle englanniksi vaan suomeksi.
- (6) SO: Tänään minä en kirjoita sinulle ruotsiksi vaan suomeksi.
- (7) OK: Me asumme nyt Hannan opiskelija-asunnossa Espoossa.
- (8) SO: Asun opiskelija-asunnossa Mariehemissä.
- (9) OK: Missä sinä asut?
- (10) SO: Missä sinä asut nyt?

- (11) OK: Nähdään, moikka!
- (12) SO: Nähdään kotona.
  
- (13) OK: Paljon terveisiä Lauralle! Terveisin Pedro
- (14) SO: Paljon terveisiä. Terveisin NN

Teksti B on tekstityypiltään kuvaus. Tehtävänannossa oppijoita pyydettiin kuvailemaan fantasiahenkilön tai oikean henkilön ulkoisia ja sisäisiä ominaisuuksia. Taulukko 1 näyttää varsinkin oppikirjan ensimmäisestä kappaleesta otettujen sanojen määrän vähentyneen melkoisesti. Huomion arvoista on, että oppikirjan kolmannen ja neljännen kappaleen sanoja on hyödynnetty lähes yhtä paljon. Tämä selittyy ainakin osin sillä, että jo kolmannessa kappaleessa tutustutaan moniin adjektiiveihin, ja niiden käyttöä harjoitellaan eri tavoin. Neljännen kappaleen sanastossa on uusia adjektiiveja ja perheeseen ja sukulaisuuteen liittyvää sanastoa. Myös omistusrakenteeseen tutustuminen on lisännyt samankaltaisuutta oppikirjan ja tuotosten sanastojen välillä.

Oppijoiden tekstissä B ja oppikirjassa on paljon identtisiä tai melkein identtisiä ilmauksia ja lauseita, kuten esimerkeistä (15–28) näkyy.

- (15) OK: Hänellä on ruskeat silmät.
- (16) SO: Bernandilla on ruskeat silmät ja hän on 178 cm pitkä --- .
  
- (17) OK: Hänellä on lyhyt tumma kihara tukka.
- (18) SO: Hänellä on lyhyt tumma tukka.
  
- (19) OK: Hänellä on siniset silmät ja silmälasit.
- (20) SO: Hänellä on vihreät silmät ja silmälasit.
  
- (21) OK: Minä olen naimisissa Veijon kanssa.
- (22) SO: Rauni on naimisissa Harrin kanssa.
  
- (23) OK: Hän on opiskelija, mutta hän on myös töissä kaupassa.
- (24) SO: Joskus hän töissä kaupassa.
  
- (25) OK: Hän on pitkä ja hoikka.
- (26) SO: Hän on melko pitkä, hoikka ja hyvännäköinen.

(27) OK: Kotona Etelä-Afrikassa minulla on iso perhe --- .

(28) SO: Annalla on iso perhe Uumajassa.

Tekstissä C oppijat kuvaavat yhden tavallisen päivänsä kulkua. Tehtävä on suoraan verrannollinen oppikirjan lukutekstin kanssa, mikä vaikuttaa juuri tästä kappaleesta otettujen sanojen korkeaan määrään. Tekstin C sanasto poikkeaa tekstin A ja B sanastoista siten, että edeltävän eli neljännen lukukappaleen sanastosta otettujen sanojen määrä on alhainen. Oletettavasti tämä johtuu siitä, että oppijat ovat fokuoituneet tuotoksissa enemmän tapahtuviin toimintoihin kuin niiden laatuun, jolloin adjektiiveja ei ole ollut tarpeen käyttää. Poikkeavaa verrattuna teksteihin A ja B on sekin, että oppikirjan kolmen ensimmäisen kappaleen sanoja on mukana hyvin tasaisesti.

Oppikirjan viidennessä kappaleessa on paljon arkielämään liittyvää sanastoa. Esimerkeistä (29–42) näkee, että oppijat ovat siirtäneet suoraan valmiiksi saamiaan malleja tuotoksiinsa.

(29) OK: Keittiössä minä teen aamupalaa.

(30) SO: Keittiössä minä teen aamupalaa.

(31) OK: Lapset syövät tavallisesti muroja ja minä syön voileipää.

(32) SO: Yleensä syön voileipää ja muna.

(33) OK: Syömme ja sitten katsomme vähän televisiota.

(34) SO: Syön päivällistä ja katson vähän televisiota.

(35) OK: Töissä kirjoitan ja piirrän paljon tietokoneella.

(36) SO: Minä luen paljon iltapäivällä ja kirjoitan tietokoneella.

(37) OK: Minä olen väsynyt --- . --- lapset menevät nukkumaan.

(38) SO: Jälkeen minä olen usein väsynyt mutta minä täytyy opiskella vähän ennen menen nukkumaan.

(39) OK: Illalla minä urheilen Kallen kanssa. Tänään uimm uimahallissa.

(40) SO: --- siellä urhellen uin uimahallissa.

(41) OK: Menen ensin suihkuun ja sitten keittiöön.

(42) SO: Menen ensin suihkuun, sitten pukeutun ja meikaan.

Oppijoiden tuotoksissa on paljon yksittäisiä sanoja ja fraaseja, jotka esiintyvät oppikirjan viidessä ensimmäisessä kappaleessa. Näyttää siltä, että oppijat ovat tutustuneet kappaleiden tekstien sanastoon. Sitä, ovatko he oppineet ja sisäistäneet sanan tai fraasin merkityksen ja käyttöympäristön, on mahdotonta sanoa. Useissa tapauksissa oppija on nostanut esimerkiksi kokonaisen fraasin suoraan oppikirjasta tekstiinsä. Toisissa tapauksissa oppijat ovat muokanneet fraasia tai lausetta hieman. Kappaleista on myös poimittu yksittäisiä sanoja ja sijoitettu omaan tuotokseen.

Tutkijoiden näkemysten mukaan oppijat käyttävät erilaisia strategioita opetellessaan sanoja ja niiden käyttöä. Muun muassa Masonen (2003: 94–104) toteaa uuden kielen oppijoiden tavoitteena olevan sellainen kielenkäyttö, jossa yksittäiset sanat yhdistyvät toisiinsa ja muodostavat ymmärrettäviä kokonaisuuksia. Hänen mukaansa oppijoilla on kuitenkin niin sanottu sanalistojen vaihe, jolloin he opettelevat yksittäisiä sanoja ja myös kääntävät ne äidinkielelleen. Masonen viittaa Nationiin (1990: 126), jonka mukaan oppiminen on tehokasta, kun oppija yhdistää uuteen sanaan äidinkiellensä sanan ja pyrkii luomaan assosiaation sanojen välille.

Puro (1999: 4–5) viittaa Lewisin (1993: vi, 19) käsitykseen, jonka mukaan rakenteena kielioppi on alisteinen leksikolle, ja suurin osa kielestä sisältyy monisanaisiin jaksoihin. Kohdekielen oppijat hyötyvät tällaisista kokonaisista sanajaksoista ja pystyvät jo opintojensa alussa tuottamaan valmiita ilmauksia. Myös Schmitt (2010: 138) on sitä mieltä, että kohdekielen omaksumisen alussa sanasekvenssit ovat nopea tapa kommunikatiivisuuteen. Tämä johtaa nopeampaan sulautumiseen samankaltaisten oppijoiden ryhmään, mikä puolestaan lisää kielellistä syötöstä. Sanasekvenssien oppiminen voi myös edistää tulevaa kielenoppimista. Schmitt ja Carter (2004: 12) toteavat kieltä omaksuttavan kahdella tavalla: sekä sanasekvensseihin että kielioppiin perustuen. Sundmanin (2011: 335) mukaan on ilmeistä, että sanasekvenssien oppimisella on keskeinen rooli kielen rakenteen oppimisessa. Suoran, eksplisiittisen, kieliopin opettamisen on nähty edistävän oppimista hyvin tehokkaalla tavalla. Sundman (2011: 335) huomauttaa kuitenkin,

että pitää olla tietoinen siitä, että kouluopetuksessa oppijat oppivat kieliopillisia muotoja ja rakenteita myös epäsuorasti, implisiittisesti. Epäsuoraan oppimiseen tarvitaan runsasta kielellistä syötöstä.

### **3.1. Tuotosten ja oppikirjan sanojen sanaluokista**

#### **3.1.1. Tuotosten sanojen sanaluokista**

Vieraan kielen sanojen sanaluokkien tunnistaminen voi olla oppijalle työlästä. Siksi pä alkeiskurssillamme sanaluokkiin perehdytään tarkasti jo aivan opintojen alussa, ja oppikirjaan kuuluvissa sanastoissa sanaluokat on myös mainittu. Tanner (2008: 6) toteaa, että “sanaluokka on erittäin tärkeä sanan hahmottaja ja sanastoon sijoittaja, ja se vaikuttaa sanan käyttöön.” Moni suomenoppija tuskailee sanojen oppimisen kanssa, koska hän ei useinkaan saa tukea esimerkiksi omasta äidinkielistään. Suomen kielen sanasto on hyvin omaperäistä, ja oppijalle sanat ovat läpinäkymättömiä. Ingo (2000: 100–101) tiivistääkin seuraavasti: “Sanaston oppiminen on näin ollen useimmille suomea toisena ja vieraana kielenä opiskeleville aikamoinen tehtävä.”

Sananluokan tuntemisesta oppija saa tietoa muun muassa siitä, miten sana taipuu, millaisia affikseja sen kanssa voi käyttää ja millaisten määritteiden kanssa se voi esiintyä. Lisäksi sanaluokka kertoo, minkä muiden sanojen kanssa sana esiintyy ja minä lauseenjäsenenä se voi toimia. (Tanner 2008: 6.)

Seuraavassa tarkastelen, millaista oppijoiden käyttämä sanasto on sanaluokkien näkökulmasta. Ensin esittelen tulokset oppijoiden tuotoksista ja sitten oppikirjan sanastosta. Vertaan myös, miten tulokset korreloivat keskenään. Sanaluokat on merkitty korpuksiin käsin. Tästä syystä kaikki saneet tuotoksista ja oppikirjan sanastosta eivät ole mukana. Jokaisesta tekstistä ja oppikirjan kappaleesta on otettu 200 saneen otokset. Saneet on kerätty frekvenssisanalistoista, jotka on saatu *WordSmith Tools* -ohjelman *WordList*-työkalulla.

**TAULUKKO 2.** *Tuotosten sanaluokat*

| Sanaluokka   | Teksti A |      | Teksti B |      | Teksti C |      | Koko aineisto |      |
|--------------|----------|------|----------|------|----------|------|---------------|------|
|              | f        | %    | f        | %    | f        | %    | f             | %    |
| substantiivi | 68       | 34,0 | 100      | 50,0 | 81       | 40,5 | 249           | 41,5 |
| adjektiivi   | 12       | 6,0  | 43       | 21,5 | 4        | 2,0  | 59            | 9,8  |
| pronomini    | 16       | 8,0  | 11       | 5,5  | 12       | 6,0  | 39            | 6,5  |
| lukusana     | 9        | 4,5  | 4        | 2,0  | 14       | 7,0  | 27            | 4,5  |
| verbi        | 51       | 25,5 | 21       | 10,5 | 53       | 26,5 | 125           | 20,8 |
| adverbi      | 44       | 22,0 | 21       | 10,5 | 36       | 18,0 | 101           | 16,8 |
| Summa        | 200      | 100  | 200      | 100  | 200      | 100  | 600           | 99,9 |

Tuotosten sanaluokkajakaumaan ovat vaikuttaneet muun muassa tekstityyppi, kirjoitusajankohta sekä oppikirjassa esiintyneet sanat ja niiden sanaluokat. Tekstissä A oppijat suosivat substantiiveja, verbejä ja adverbejä. Adverbien runsaus johtuu siitä, että oppijat käyttävät erilaisia ajanilmauksia kertoessaan, mitä he milloinkin tekevät. Koska tehtävän ajankohta kurssilla oli niin aikainen, oppijoilla ei ollut käytettävissään kovin suurta sanavarastoa, mistä johtuen heidän lauseensa olivat yksinkertaisia ja lyhyitä kuten esimerkeistä (43–46) käy ilmi. Myös tehtävän luonne vaikuttanee ilmauksen mittaan.

- (43) Terve! Mitä sinulle kuulu? Minulle kuulu hyvää!  
 (44) Missä sinä asut?  
 (45) Aamulla se on pakkasta.  
 (46) Miten menee tänään?

Tekstissä B substantiivit ja adjektiivit dominoivat. Näitä sanaluokkia on käytetty yhteensä yli 70 % kaikista sanaluokista. Tehtävän tekstityyppi on jonkun henkilön kuvaus, mikä selittää etenkin adjektiivien korkean määrän verrattuna tekstiin A ja varsinkin tekstiin C. Verbejä on paljon vähemmän kuin teksteissä A ja C. Verbien käyttö tuotoksissa on yksipuolista ja *olla*-verbi on ylikorostunut. Useimmat lauseet ovat tyyppiä *hän on* ja omistusrakenteita *hänellä on*.

- (47) Hänen tukka on iso, niinko vänäläinen karvalakki.
- (48) Nean silmät on kaunis vihreät väri.
- (49) Hänellä on lyhyt ruskea hiukset ja suuri nenä ja maha on myös iso.

Teksti C on myös substantiivi- ja verbivoittoinen. Verbejä on käytetty eniten juuri tässä tehtävässä. Oppikirjan viidennessä kappaleessa tutustutaan viiteen verbityyppiin ja niiden taivutukseen. Oppijat saavat koko joukon uusia verbejä ja voivat nyt paremmin kertoa, mitä he itse tekevät tai mitä joku muu tekee. Tästä huolimatta substantiivien määrä on yhä korkea. Tekstissä C lukusanoja on käytetty enemmän kuin edellisissä teksteissä. Syy on siinä, että oppijat hyödyntävät tarkkoja kellonaikoja kertoessaan, mitä he tekevät.

- (50) Minä herään kello 8 kun kännykän alarmi soi.
- (51) Illalla olen suomen kurssilla, kurssi alkaa kello kuusi ja päättyy kello puoli yhdeksän.
- (52) Mutta haluan työpaikassa kahdeksalta aamulla olla.

Taulukon 2 mukaan tuotosten sanoista suurin osa (41,5 %) on substantiiveja. Seuraavaksi eniten (20,8 %) teksteissä on verbejä ja sitten adverbeja (16,8 %). Adjektiivit näyttävät selvästi olevan sidoksissa tehtävän tekstityyppiin, erityyppisiin kuvauksiin ja luonnehdintoihin, koska ainoastaan 9,8 % tuotoksissa esiintyneistä sanoista on adjektiiveja. Voi olla niin, että aloittelevilla suomenoppijoilla ei ole vielä kykyä “maustaa” tuotoksiaan, vaan he keskittyvät tuottamaan pelkistettyä tekstiä, joka sekin on vaativaa. Pronominien käyttö tässä kirjoitusvaiheessa on aika niukkaa; aineistossa niitä on vain 6,5 %. Lienee niin, että vaikka pronominia olisi voinutkin käyttää, oppija on sen sijasta valinnut substantiivin. Lukusanoja aineistossa on vähiten (4,5 %), ja useimmiten ne esiintyvät ilmaistaessa kellonaikoja. Tiivistäen voi todeta, että ruotsinkielisen alkeisoppijan tuottama suomi on luonteeltaan etupäässä substantiivi-verbi-kieltä.

Halme (2009: 71–100) on tutkinut muun muassa sitä, millaisia sanaluokkajakaumia on löydettävissä sanojen tuottamistehtävistä, joiden tekemiseen osallistui kirjoitus- ja lukutaidottomia maahanmuuttajia.



Verrokkeina olivat luku- ja kirjoitustaitoiset maahanmuuttajat. Molemmat ryhmät opiskelivat suomea vieraana kielenä. Halmeen (2009: 84–86) tulosten mukaan luku- ja kirjoitustaidottomien tehtävät sisälsivät eniten substantiiveja ja sitten verbejä. Lukusanoja, adjektiiveja ja partikkeleita oli tuotettu vai yksi kutakin ja pronomineja ei ollenkaan. Verrokkien sanaluokkajakautuma oli seuraava: substantiivit, verbit, pronominit, partikkelit, adjektiivit ja lukusanat. Tuloksesta Halme (2009: 84) päättelee, että partikkelit ja pronominit ilmestyvät osaksi sanavarastoa vasta sitten, kun perusasiat osataan ilmaista. Lukusanoista hän toteaa niiden olevan alkuvaiheessa tärkeä osa sanastoa. Kieltä opittaessa enemmän lukusanojen käyttö näyttää vähenevän.

### 3.1.2. Oppikirjan sanojen sanaluokista

Suomi toisena ja vieraana kielenä -opetuksessa käytetyistä oppikirjoista on tutkittu muun muassa sanaluokkien frekvenssejä (ks. esim. Autere-Kesti 1996; Puro 1999). Kuten edellä on jo mainittu, oppikirjalla on tärkeä, jopa auktoriteetin rooli uuden kielen oppimisessa, ja oppikirjan sanasto heijastuu melkein pä korostetusti ainakin oppijoiden kirjoitustussa kielessä.

Tässä osiossa tarkastelen, miten sanaluokat jakautuvat oppikirjan “Suomen mestari 1” kolmannessa, neljännessä ja viidennessä kappaleessa. Laskelmat perustuvat 200 saneen otoksiin, jotka on kerätty frekvenssisanalistoista samoin kuin oppijoidenkin 200 saneen otokset (ks. 3.1.1). Tulokset on katsottava suuntaa antaviksi, koska oppikirjan kaikki sanat eivät ole tarkastelun kohteena.

Oppikirjan kolmannessa kappaleessa substantiivien määrä on korkea (58 %). Kappale on ensimmäinen, jossa lukutekstinä on juoksevaa tekstiä. Siten virkkeet ovat pitempiä kuin ensimmäisessä ja toisessa kappaleessa, jotka pääosin ovat dialogin muodossa. Kappaleessa oppija tutustutetaan moniin uusiin asioihin, joissa substantiivit ovat tärkeänä osana ilmaisuja. Seuraavaksi eniten kappaleessa esiintyy verbejä (21 %).

**TAULUKKO 3.** *Oppikirjan kappaleiden sanaluokat*

| Sanaluokka   | Kappale 3 |      | Kappale 4 |      | Kappale 5 |      | Summa |      |
|--------------|-----------|------|-----------|------|-----------|------|-------|------|
|              | f         | %    | f         | %    | f         | %    | f     | %    |
| substantiivi | 116       | 58,0 | 135       | 67,5 | 78        | 39,0 | 329   | 54,8 |
| adjektiivi   | 22        | 11,0 | 29        | 14,5 | 7         | 3,5  | 58    | 9,7  |
| pronomini    | 8         | 4,0  | 10        | 5,0  | 5         | 2,5  | 23    | 3,8  |
| lukusana     | 0         | 0    | 1         | 0,5  | 7         | 3,5  | 8     | 1,3  |
| verbi        | 42        | 21,0 | 16        | 8,0  | 92        | 46,0 | 150   | 25,0 |
| adverbi      | 12        | 6,0  | 9         | 4,5  | 11        | 5,5  | 32    | 5,3  |
| Summa        | 200       | 100  | 200       | 100  | 200       | 100  | 600   | 100  |

Mainittakoon vertailun vuoksi, että Autere-Kestin oppikirjatutkimuksessa (Karppinen 1996: 213) verbit ovat niin ikään toiseksi suurin sanaluokkaryhmä (23,2 %) mutta substantiivit vasta kolmanneksi suurin (22,3 %). Puron (1999: 10–11) saamien tulosten mukaan verbejä on eniten, noin kolmasosa hänen aineistonsa sanoista, ja substantiivit ovat toisella tilalla. Adjektiiveja kolmannessa kappaleessa esiintyy jonkin verran samoin kuin pronomineja ja adverbejä. Tässä 200 sanan otoksessa ei ollut ainoatakaan lukusanaa.

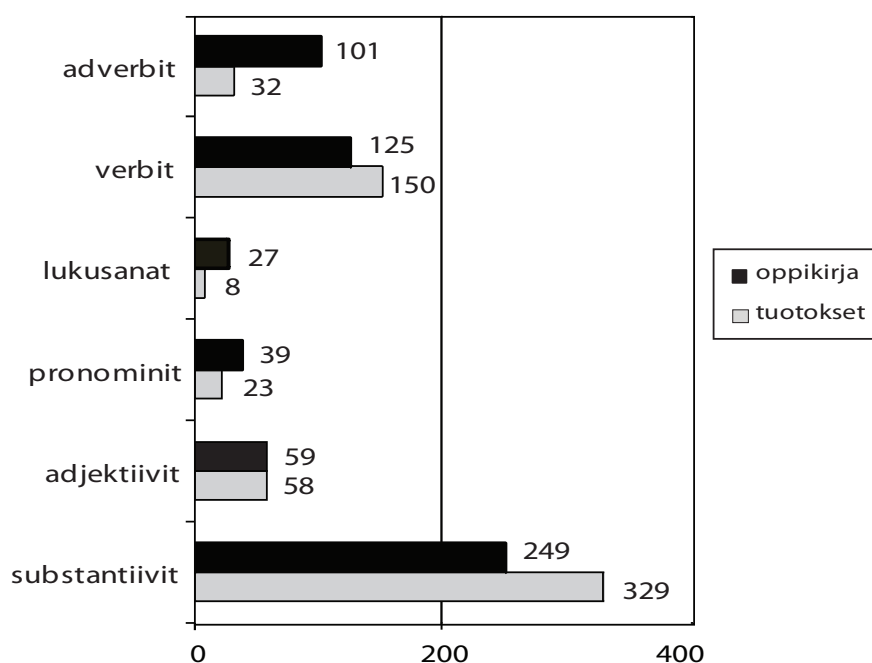
Oppikirjan neljäs kappale seuraa samaa mallia kuin kolmas kappale. Substantiivien määrä on entisestään noussut (67,5 %), kun taas verbien määrä on laskenut (8 %). Syynä lienee tekstin sisältö: eri henkilöt kertovat, keitä he ovat, minkälaisia he ovat, mitä he tekevät ja mitä heillä on tai ei ole. Dominoivana verbinä on *olla*-verbi. Adjektiiveja on esitelty useampia (14,5 %), mutta pronomineista, lukusanoista ja adverbeistä on edelleenkin mukana vain joitakin. Autere-Kestin (Karppinen 1996: 213) aineistossa adjektiivien osuus on 9 % ja Puron (1999: 11) aineistossa 6,1 %.

Oppikirjan viides kappale poikkeaa edellisistä: verbit ovat yleisempiä kuin substantiivit. Verbejä on 46 % ja substantiiveja 39 %. Tulos on odotuksenmukainen, sillä kappaleessa esitellään viiden verbityypin verbejä ja niiden taivutusta. Adjektiivien määrä on laskenut melkoisesti,

niitä on vain 7 (3,5 %). Tässä kappaleessa myös pronomineja, lukusanoja ja adverbeja on niukasti. Kokoavasti sanaluokkien yleisyysjärjestys aineistossa on seuraava: substantiivit (54,8 %), verbit (25 %), adjektiivit (9,7 %), adverbit (5,3 %), pronominit (3,8 %) ja lukusanat (1,3 %).

### 3.1.3. Tuotosten ja oppikirjan sanaluokkien vertailua

Verrattaessa keskenään oppijoiden käyttämien sanojen ja oppikirjassa esiteltyjen sanojen sanaluokkia kuvioista 1 näkee, kuinka sanaluokat tuotoksissa ja oppikirjassa suurimmaksi osaksi myötäilevät toisiaan.



**KUVIO 1.** Sanaluokat tuotoksissa ja oppikirjassa

Kuvion 1 mukaan molemmissa korpuksissa substantiivit ovat ylliedustettuina. Substantiivit on helpompi oppia kuin verbit, sillä substantiivien merkityksien havaitseminen ja ymmärtäminen on vaivattomampaa kuin verbien; tämä oppimisjärjestys on monessa suhteessa universaali (Hatch 1983: 63). Oppijoiden tuotoksissa on 80 substantiivia enemmän kuin oppikirjassa, mikä kertoo, että oppijoiden tekstit ovat rakentuneet enimmäkseen substantiiveista muiden sanaluokkien kustannuksella.

Myös verbejä esiintyy jonkin verran enemmän oppijoiden tuotoksissa kuin oppikirjassa. Syynä lienee *olla*-verbin frekventti käyttö. Adjektiivien määrä korpuksissa on kutakuinkin sama. Sitä vastoin lukusanoja ja pronomineja oppijoiden teksteissä on vähemmän kuin oppikirjassa. Lukusanojen esiintyminen lienee kytköksissä tehtävän aihepiiriin. Lukusanat eivät ole kovin suurtaajuisia oppikirjoissakaan. Autere-Kestin (Karpinen 1996: 213) oppikirjojen dialogiaineistossa lukusanojen osuus on 2,8 % ja vastaavasti Puron (1999: 10–11) oppikirja-aineistossa noin 2 %.

Pronominien vähäisempään käyttöön kuin oppikirjassa voinee syynä olla se, että oppijat käyttävät niiden sijasta mieluummin substantiiveja. “Suomen mestari 1:ssä” pronomineja on keskimäärin 3,8 % (taulukko 3). Oppikirjassa pronomineista otetaan esille vain persoonapronominit ja demonstratiivipronominit.

Oppijoiden teksteissä adverbeja on huomattavasti vähemmän kuin oppikirjassa. On luultavaa, että opintojensa alkuvaiheessa oleva oppija ei vielä kykene käyttämään lauseissaan ja virkkeissään mitään ylimääräistä ainesta, vaan hän keskittyy tuottamaan vain yksinkertaista ja pelkistettyä kieltä. Oppikirjasta “Suomen mestari 1” tarkastelluissa kolmessa luku-kappaleessa adverbien määrä on 5,3 % (taulukko 3).

Oppijoiden tuotosten ja oppikirjan sanojen sanaluokkien vertailusta käy ilmi, että molemmissa korpuksissa käytetään paljon samoihin sanaluokkiin kuuluvia sanoja. Sanaluokkien yleisyysjärjestys korpuksissa on seuraava:

- oppijoiden tuotokset:
  1. substantiivit, 2. verbit, 3. adjektiivit, 4. adverbit, 5. pronominit, 6. lukusanat
- oppikirjan sanasto:
  1. substantiivit, 2. verbit, 3. adverbit, 4. adjektiivit, 5. pronominit, 6. lukusanat

Suomen yleiskielessä leksikaalinen sanaluokkajakauma on Taajuus-sanaston (Saukkonen ym. 1979: 11) mukaan seuraava:

1. substantiivit, 2. adjektiivit, 3. verbit, 4. adverbit, 5. lukusanat, 6. pronominit.

Verrattaessa toisiinsa oppijoiden käyttämien ja suomen yleiskielen sanaluokkien suosittuusjärjestystä nähdään, että järjestykset eroavat muiden sanaluokkien paitsi substantiivien kohdalla. Eron lienee useitakin syitä. Oppijoiden tuotosten korpus on pieni, ja näin ollen tulokset on katsottava vain suuntaa antaviksi. Myös se, että tuotokset edustavat tiettyjä tekstityyppejä, vaikuttanee jonkin sanaluokan sanojen taajempaan tai vähäisempään käyttöön.

### 3.2. Tuotosten yleisimmistä sanoista

Jotta saisi konkreettisemmän kuvan oppijoiden käyttämästä sanastosta ja käytettyjen sanojen yleisyydestä tuotoksissa, korpuksista on *Word-List*-ohjelmalla otettu saatu sanojen frekvenssit. Taulukkoon 4 on kerätty kustakin tuotoksesta 25 yleisintä sanaa. Sanat ovat siinä muodossa kuin ne ovat esiintyneet oppijoiden teksteissä. Tästä johtuen sama sana voi toistua eri muodoissa.

Tekstin A eniten käytetty sana on konjunktio *ja*. Tämä osoittaa sen, että oppijat ovat kyenneet rinnastamaan ainakin kaksi yksittäistä sanaa tai kaksi lausetta toisiinsa jo opintojen alkuvaiheessa. Tuotettuun tekstiin ei siis pelkästään sisälly yksittäisiä, toisiaan seuraavia päälauseita. *Ja*-konjunktin lisäksi taulukossa on mukana alistuskonjunktio *kun* ja rinnastuskonjunktio *mutta*. Muista adverbeista mainittakoon *sitten*, joka toistuu tekstissä usein. Oppijat käyttävät adverbia siirtyessään tekstissään yhdestä asiasta toiseen.

Verbeistä *olla*-verbi ja sen yksikön 3. persoonan muoto on eniten käytetty verbi tekstissä A. Muita yleisiä verbejä ovat *mennä*, *syödä* ja *kuulua*, joka esiintyy fraaseissa *Mitä sinulle kuuluu?* ja *Minulle kuuluu hyvää*. *Mennä*-verbin yleisyys johtuu siitä, että oppijat kirjoittavat sähköpostiviestissään, kuinka he menevät illalla suomen kurssille. Kieltoverbin *ei* käyttö on niukanlaista, ja se esiintyy vain yksikön 1. persoonan muodossa. Tähän lienee syynä tuotoksen tekstityyppi, sähköpostiviesti, ja se, että oppikirjan kolmen ensimmäisen kappaleen lukuteksteissä negatiivisia lauseita on yhteensä vain seitsemän. Oppijalta saattaa näin ollen puuttua tarvitsemansa negatiivisen lauseen malli.

**TAULUKKO 4.** *Tuotosten 25 yleisintä sanaa teksteittäin*

| Sija | Teksti A | f   | Teksti B | f   | Teksti C  | f   |
|------|----------|-----|----------|-----|-----------|-----|
| 1    | ja       | 224 | on       | 251 | ja        | 138 |
| 2    | on       | 167 | ja       | 145 | minä      | 117 |
| 3    | minä     | 164 | hän      | 107 | kello     | 70  |
| 4    | kello    | 77  | hänellä  | 61  | on        | 66  |
| 5    | minun    | 75  | hänen    | 31  | menen     | 54  |
| 6    | olen     | 69  | vuotta   | 20  | minun     | 54  |
| 7    | menen    | 56  | tukka    | 18  | syön      | 48  |
| 8    | syön     | 48  | ei       | 17  | sitten    | 38  |
| 9    | sitten   | 45  | myös     | 17  | kanssa    | 36  |
| 10   | se       | 43  | ole      | 17  | olen      | 36  |
| 11   | kanssa   | 41  | silmät   | 17  | kun       | 28  |
| 12   | me       | 34  | vanha    | 17  | illalla   | 24  |
| 13   | illalla  | 33  | mutta    | 16  | minulla   | 24  |
| 14   | kun      | 32  | lyhyt    | 15  | päivä     | 23  |
| 15   | mutta    | 30  | asuu     | 14  | joskus    | 20  |
| 16   | minulla  | 29  | he       | 13  | se        | 20  |
| 17   | nyt      | 29  | kanssa   | 13  | me        | 19  |
| 18   | päivä    | 29  | nimi     | 13  | mutta     | 18  |
| 19   | joskus   | 24  | pitkä    | 13  | aamupalaa | 17  |
| 20   | tänään   | 24  | perhe    | 12  | yleensä   | 17  |
| 21   | kuuluu   | 22  | ovat     | 11  | luen      | 16  |
| 22   | paljon   | 22  | kaksi    | 10  | nukkumaan | 15  |
| 23   | suomen   | 21  | musta    | 10  | teen      | 15  |
| 24   | kotona   | 20  | heillä   | 9   | kotiin    | 14  |
| 25   | en       | 19  | iso      | 9   | en        | 13  |

Otoksen 25 sanan joukossa on vain viisi verbiä, ja ne on taivutettu *minä*-muodossa. On yleistä, että verbien määrä on alkeiskurssin alussa pieni, substantiivien kustannuksella, ja enin osa asioista ilmaistaan *olla*-verbin välityksellä. Siitonen ja Niemelä (2011: 242–279) ovat tutkineet edistyneiden suomenoppijoiden tenttivastauksia, ja yhtenä tutkimuksen kohteena on *olla*-verbin käyttö. Siitosen ja Niemelän (2011: 262–263) aineistoon pohjautuvan laskelmani mukaan *olla*-verbiä esiintyi predikaattina kaikista predikaattiverbeistä heidän tutkimuksensa alussa keskimäärin 37,4 %:isesti ja lopussa 39,5 %:isesti Odotuksenvastaisesti *olla*-verbi on siis melko yleinen vielä edistyneilläkin oppijoilla.

Pronomineista tekstissä A ovat mukana ainoastaan *minä*, *me* ja *se*. Nominatiivisen *minä*-muodon lisäksi listalla on pronominin genetiivimuoto *minun* ja adessiivimuoto *minulla*. Adjektiiveja ja lukusanoja ei ole ollenkaan 25 sanan joukossa.

Tekstin B frekventein sana on *olla*-verbi taivutettuna yksikön 3. persoonassa. Käyttötaajuus tässä tekstissä, joka on henkilökuvaus, on suurelta osin kytköksissä omistusrakenteeseen, kun oppijat kertovat, mitä heidän henkilöillään on tai ei ole. Kymmenentenä listalla onkin *olla*-verbin kielto muoto *ole*. Monikon 3. persoonan muotoa *ovat* esiintyy tekstissä jonkin verran. *Olla*-verbin lisäksi tekstissä B esiintyy verbi *asua* ja kieltoverbi 3. persoonan muodossa *ei*.

Tekstissä B oli tarkoitus harjoitella adjektiivien käyttöä. Näyttää siltä, että adjektiivien kirjo on niin jakautunutta, että suurtaajuisten sanojen listalle ylsivät vain adjektiivit *vanha*, *lyhyt*, *pitkä*, *musta* ja *iso*. Pronomineista ovat mukana vain *hän*, genetiivi *hänen* ja adessiivi *hänellä* sekä *he* ja adessiivi *heillä*. Pronominien adessiivimuodot *hänellä* ja *heillä* esiintyvät omistusrakenteissa. Adverbeja tekstissä B on vain muutama, ja niistä suurin osa on *ja*-rinnastuskonjunktioita. Lukusanoista on vain yksi esiintymä, *kaksi*.

Tekstin C yleisin sana on rinnastuskonjunktio *ja*. Muista konjunktioista on mukana *kun* ja *mutta*. Poikkeavaa verrattuna tekstiin A ja B on se, että *olla*-verbi on vasta sijalla neljä. Verbistä esiintyy vain muodot *on* ja *olen*. Verbien määrä ei ole lisääntynyt jostain syystä, vaikka oppi-



kirjan kappaleessa viisi esitellään paljon eri tyyppeihin kuuluvia verbejä. Tekstin muita verbejä ovat *mennä*, *syödä*, *lukea* ja *tehdä*, ja ne kaikki esiintyvät yksikön 1. persoonan muodossa. Listalla on myös verbi *nukkua* 3. infinitiivin illatiivimuodossa, *nukkumaan*. Alkeistason opiskelijat eivät tunne tätä muotoa, eikä sitä ole opetuksessa heille selitetykään. Oppikirjassa muotoa on käytetty pari kertaa. Kieltoverbillisiä lauseita teksteissä on vähän, ja yleisin muoto on *minä*-persoonan *en*.

Tekstissä C käytetään samoja pronomineja kuin tekstissä A; tekstissä B on käytetty pronomineja *minä*, *minun*, *minulla*, *se* ja *me*. Adjektiiveja ja lukusanoja ei ole ollenkaan 25 sanan joukossa.

Oppijoiden sanasto ei ole kovin vaihtelevaa taulukon 4 mukaan. Samat sanat ja sanamuodot toistuvat tekstistä toiseen. Tekstit ovat hyvin *minä*-keskeisiä, poikkeuksena teksti B, jossa persoonana on *hän*. Tähän ovat vaikuttaneet tehtävänanto ja tehtävän tekstityyppi. Siksi olisikin äärimmäisen tärkeää, että L2-oppijoille annettaisiin vaihtelevia kirjoitustehtäviä ja opetettaisiin sanastoa myös oppikirjan ulkopuolelta.

Taulukon 4 pohjalta olen koostanut taulukkoon 5 kaikkien kolmen tekstien 25 yleisintä sanaa ja niiden absoluuttiset määrät. Edelleen sanat ovat siinä muodossa kuin ne ovat esiintyneet teksteissä.

Verbi *olla* on koko korpuksen frekventein sana. Se esiintyy ainoastaan muodoissa *on* ja *olen* sekä kieltomuodossa *ole*. Muita verbejä otoksessa ovat *mennä* ja *syödä*. Kieltoauseita korpuksessa ei kovinkaan usein esiinny, ja kieltoverbiä edustaa vain 3. persoonan muoto *ei*. Toiseksi yleisin sana on rinnastuskonjunktio *ja*. Korpuksen muut konjunktiot ovat *mutta* ja *kun*.

Pronominit *minä* ja *hän* ovat yleisiä. Niistä on listalla mukana myös genetiivimuodot *minun*, *hänen* ja adessiivimuodot *minulla*, *hänellä*. Muita pronomineja ovat nominatiivimuotoiset *se* ja *me*.

Adjektiivit ja lukusanat ovat jääneet otoksen ulkopuolelle. Partikleista yllä mainittujen konjunktoiden lisäksi esiintyvät *kanssa*, *sitten*, *nyt*, *myös* ja *paljon*. Mielenkiintoinen havainto on se, että vaikka substantiivit ovat korpuksen dominoiva sanaluokka, niitä on 25 sanan otoksessa vain kolme, nimittäin *kello*, *illalla* ja *päivä*.

**TAULUKKO 5.** *Tuotosten 25 yleisintä sanaa koko korpuksessa*

| Sija | Tekstit A + B + C | f   |
|------|-------------------|-----|
| 1    | on                | 418 |
| 2    | ja                | 369 |
| 3    | minä              | 164 |
| 4    | hän               | 121 |
| 5    | minun             | 79  |
| 6    | kello             | 78  |
| 7    | olen              | 69  |
| 8    | hänellä           | 62  |
| 9    | menen             | 56  |
| 10   | kanssa            | 54  |
| 11   | syön              | 48  |
| 12   | se                | 47  |
| 13   | mutta             | 46  |
| 14   | sitten            | 45  |
| 15   | illalla           | 37  |
| 16   | kun               | 36  |
| 17   | hänen             | 34  |
| 18   | me                | 34  |
| 19   | nyt               | 31  |
| 20   | ei                | 30  |
| 21   | myös              | 30  |
| 22   | minulla           | 29  |
| 23   | päivä             | 29  |
| 24   | ole               | 28  |
| 25   | paljon            | 27  |

On oletettavaa, että oppijansuomen sanasto eroaa suomen yleiskielen sanastosta. Taulukossa 6 esittelen korpukseni 25 yleisintä lekseemiä ja vertaan niitä Taajuussanaston (Saukkonen ym. 1979: 41) lekseemeihin.

**TAULUKKO 6.** *Tuotosten 25 yleisimmän lekseemin ja Taajuussanaston vertailua*

| <b>Sija</b> | <b>Tekstit A + B + C</b> | <b>Taajuussanasto</b> |
|-------------|--------------------------|-----------------------|
| 1           | <i>olla</i>              | <i>olla</i>           |
| 2           | <i>ja</i>                | <i>ja</i>             |
| 3           | <i>minä</i>              | <i>se</i>             |
| 4           | <i>hän</i>               | <i>ei</i>             |
| 5           | <i>kello</i>             | <i>joka</i>           |
| 6           | <i>mennä</i>             | <i>että</i>           |
| 7           | <i>kanssa</i>            | <i>tämä</i>           |
| 8           | <i>syödä</i>             | <i>hän</i>            |
| 9           | <i>se</i>                | <i>voida</i>          |
| 10          | <i>mutta</i>             | <i>saada</i>          |
| 11          | <i>sitten</i>            | <i>mutta</i>          |
| 12          | <i>ilta</i>              | <i>niin</i>           |
| 13          | <i>kun</i>               | <i>kuin</i>           |
| 14          | <i>me</i>                | <i>ne</i>             |
| 15          | <i>nyt</i>               | <i>kun</i>            |
| 16          | <i>ei</i>                | <i>tulla</i>          |
| 17          | <i>myös</i>              | <i>myös</i>           |
| 18          | <i>päivä</i>             | <i>tai</i>            |
| 19          | <i>paljon</i>            | <i>kaikki</i>         |
| 20          | <i>tänään</i>            | <i>aika</i>           |
| 21          | <i>joskus</i>            | <i>me</i>             |
| 22          | <i>vuosi</i>             | <i>suuri</i>          |
| 23          | <i>kaksi</i>             | <i>vuosi</i>          |
| 24          | <i>kuulua</i>            | <i>mikä</i>           |
| 25          | <i>suomi</i>             | <i>toinen</i>         |

Taulukossa 6 olen kursivoinut ne sanat, jotka esiintyvät sekä oppijoiden tuotoksissa että Taajuussanastossa. Sanojen joukossa on kymmenen sanaa, jotka ovat samoja: verbi *olla* ja kieltoverbi, partikkelit *ja*, *mutta* ja *kun*, *myös*, pronominit *se*, *hän* ja *me* ja ainoana substantiivina *vuosi*.

Tulos on jonkin verran suuntaa antava. Oppijankieli on *minä*-keskeistä ja verbien variaatio on niukka. Konjunktioista käytetään aktiivisesti vain muutamaa, ja muitakin partikkeleita on vain satunnaisesti. Kuten jo aikaisemmin on todettu, substantiivit, adjektiivit ja lukusanat eivät juuri ole edustettuina 25 yleisimmän sanan otoksessa.

## 5. Lopuksi

Artikkelissani olen pohtinut, onko suomen kielen alkeiskurssilla käytetyn oppikirjan sanasto vaikuttanut ruotsinkielisten suomenoppijoiden kirjallisten tuotosten kieleen. Verrattaessa keskenään yhtäältä tuotosten, toisaalta oppikirjojen sanoja ja sanamuotoja käy ilmi, että keskimäärin 34 % oppijoiden tuotosten yksittäisistä sanoista, monisanaisista sanasekvensseistä ja kokonaisista lauseista on identtisiä oppikirjan kanssa. On kuitenkin mahdotonta sanoa, ovatko oppijat todella oppineet käyttämänsä sanat ja niiden käyttöympäristöt. Nykyiset korpukseni ovat pieniä, joten tarkempien tulosten saamiseksi aineistoa on kasvatettava; myös tutkimusmetodeja olisi kehitettävä.

Tuotosten sanaluokkien selvittäminen paljastaa, että substantiivit ja verbit ovat yleisimmät sanaluokat. Aloittelevan oppijan sanavarasto on suppea, ja myös tuotoksen tekstityyppi vaikuttanee eri sanaluokkien käyttöön. Tuotoksissa esiintyy kaikkien sanaluokkien sanoja, mutta niiden frekvenssit eivät ole kovin korkeita.

Tarkasteltaessa tuotosten yleisimpien sanojen esiintymiä nähdään, että rinnastuskonjunktio *ja* ja *olla*-verbi ovat frekventeimmät. Muiden konjunktoiden ja varsinkin verbien käyttö on hyvin vähäistä. Syynä lienee tuotosten kirjoitusajankohdat: ne kirjoitettiin vain muutamien opintoviikkojen jälkeen. Pronominien *minä* ja *hän* asema korostuu. Teksteistä kaksi on aiheiltaan *minä*-keskeisiä, ja yhtä teemaa useimmat

kirjoittajat lähestyvät *hän*-perspektiivistä. Adverbit ja substantiivit eivät ole yleisyydeltään niin tavallisia kuin voisi luulla, ja adjektiivit jäävät tutkimuksessani kokonaan ulkopuolelle. Lukusanoista on mukana vain yksi.

On ilmeistä, että sanaston tutkimukseen, opettamiseen, oppimiseen ja omaksumiseen tulisi kiinnittää paljon nykyistä enemmän huomiota. Relevanttien tutkimusmenetelmien kehittäminen sanastotutkimuksen alalle on mitä tärkeintä. Sanat ovat hyvän kommunikoinnin peruselementtejä, ja siksi niistä pitäisi saada tutkimuksen kautta niin paljon informaatiota kuin mahdollista.

## Lähteet

- Aronen, Juha-Matti, Margit Kuusk 2010. Sanalla sanoen – ühe sõnaga? – Lähivördlusi. *Lähivertailuja* 19, 106–119. <http://dx.doi.org/10.5128/LV19.07>
- Autere-Kesti, Anni 1996. Kuvaus eräiden suomi vieraana kielenä -oppikirjojen dialogeista. Pro gradu -työ. Oulun yliopiston suomen ja saamen kielen ja logopedian laitos.
- Conklin, Cathy, Norbert Schmitt 2008. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? – *Applied Linguistics* 29 (1), 72–89. <http://dx.doi.org/10.1093/applin/amm022>
- Gehring, Sonja, Sanni Heinzmann 2010. Suomen mestari 1. Suomen kielen oppikirja aikuisille. Helsinki: Oy Finn Lectura Ab.
- Grönholm, Maija 1993. Tempusten ja modusten käyttö toisen kielen kirjoittamisessa. – *Virittäjä* 97 (3), 410–420.
- Halme, Niina 2009. Luku- ja kirjoitustaidoton mutta ei sanaton. Aikuisten luku- ja kirjoitustaidottomien maahanmuuttajien luku- ja kirjoitustaito L2-kielen alkeistasolla. – Anu Rekimies, Kirsti Siitonen (Toim.). *Suunnaksi suomen kieli*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 80. Turku: Uniprint, 71–100.
- Hatch, Evelyn Marcussen 1983. *Psycholinguistics: A Second Language Perspectives*. Los Angeles: University of California.
- Ingo, Rune 2000. Suomen kieli vieraan silmin. Vaasan yliopiston käännösteorian ja ammattikielten tutkijaryhmän julkaisut 26. Vaasa: Vaasan yliopisto.

- Jantunen, Jarmo Harri 2011. Kansainvälinen oppijansuomen korpus (ICLFI): typologia, taustamuuttajat ja annotointi. – *Lähivörtlusi. Lähivertailuja* 21, 86–105. <http://dx.doi.org/10.5128/LV21.04>
- Jantunen, Jarmo Harri, Saana Piltonen 2009. Oppijansuomen ja viron sähköiset tutkimusaineistot. – *Virittäjä* 113 (3), 449–458.
- Järventausta, Marja 2004. Sanakirjat suomenoppijan apuna. Suomen ja saamen kielen ja logopedian laitoksen julkaisuja 24. Oulu: Oulun yliopisto, 24–44.
- Karppinen, Kaisa 1996. Suomi toisena ja vieraana kielenä – oppikirjojen dialogien kuvausta. – Helena Sulkala, Minna Zaman-Zadeh (Toim.). Suomenoppijan kieli tutkimuskohteena. Suomen ja saamen kielen ja logopedian laitoksen julkaisuja 5. Oulu: Oulun yliopisto, 209–220.
- Krapels, Alexandra Rowe 1990. An overview of second language writing process research. – Barbara Kroll (Ed.). *Second Language Writing. Research Insights for the Classroom*. Cambridge: Cambridge University Press, 37–56.
- Lewis, Michael 1993. *The Lexical Approach. The State of ELT and a Way forward*. London: Language Teaching Publications.
- Li, Jie, Norbert Schmitt 2009. The aquisition of lexical phrases in academic writing. A longitudinal case study. – *Journal of Second Language Writing* 18 (2), 85–102. <http://dx.doi.org/10.1016/j.jslw.2009.02.001>
- Lundberg, Gun 2011. *De första årens engelska*. Lund: Studentlitteratur AB.
- Masonen, Virpi 2003. Sanasto, opettaja ja kielenoppija. – Marjo Mela, Pirjo Mikkonen (Toim.). *Suomi kakkonen. Opas opettajille*. Tietolipas 198. Helsinki: Suomalaisen Kirjallisuuden Seura, 93–105.
- Meara, Paul 2009. *Connected Words. Word Associations and Second Language Vocabulary Acquisition*. *Language Learning & Language Teaching* 24. Amsterdam: John Benjamins Publishing Company.
- Milton, James 2009. *Measuring Second Language Vocabulary Acquisition*. SLA. *Second Language Acquisition*. Bristol: Multilingual Matters.
- Määttä, Tuija 2010a. “Haluan luen kirjan ja opiskellan piene Suomi”. *Kansainvälinen suomen tutkimushanke*. – *Liekki* 3, 16–17.
- Määttä, Tuija 2010b. “Ostaisin lahjaan sairaita lapsia”. *Oppijankielen tutkimushanke*. – *Sutina* 1, 18–20.
- Nation, I. S. P. 1990. *Teaching and Learning Vocabulary*. New York: Newbury House Publishers.
- Puro, Tarja 1999. Sanastollinen tieto ja suomen kielen oppikirjojen sanasto. – *Virittäjä* 103 (1), 2–26.
- Saukkonen, Pauli, Marjatta Haipus, Antero Niemikorpi, Helena Sulkala 1979. *Suomen kielen taajuussanasto*. WSOY: Porvoo–Helsinki.

- Schmitt, Norbert 2000. *Vocabulary in Language Teaching*. Cambridge Language Education. Cambridge: Cambridge University Press.
- Schmitt, Norbert 2010. *Researching Vocabulary. A Vocabulary Research Manual*. – *Research and Practice in Applied Linguistics*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Schmitt, Norbert, Ronald Carter 2000. Lexical phrases in language learning. – *The Language Teacher* 24 (8), 6–10.
- Schmitt, Norbert, Ronald Carter 2004. Formulaic sequences in action: An introduction. – Schmitt, Norbert (Ed.). *Formulaic Sequences: Acquisition, Processing, and Use*. *Language Learning & Language Teaching* 9. John Benjamins, 1–22.
- Schoonen, Rob, Patrick Snellings, Marie Stevenson, Amos van Geldern 2009. Towards a blueprint of the foreign language writer: The linguistic and cognitive demand of foreign language writing. – Rosa M. Manchon (Ed.). *Writing in Foreign Language Contexts. Learning, Teaching and Research*. Bristol: Multilingual Matters, 77–101.
- Siitonen, Kirsti, Jenny Niemelä 2011. Mitä pitkittäistutkimus paljastaa edistyneiden suomenoppijoiden kielitaidosta? – *Lähivördlusi. Lähivertailuja* 21, 242–279. <http://dx.doi.org/10.5128/LV21.10>
- Sundman, Marketta 2011. Helsekvenser i finska grundskoleelevers skrivande. – Ann-Catrine Edlund, Ingmarie Mellenius (Red.). *Svenskans beskrivning 31. Förhandlingar vid Trettioförsta sammankomsten för svenskans beskrivning*. Umeå den 20–21 maj 2010. Umeå: Umeå universitet, 327–336.
- Tanner, Johanna 2008. S2-oppikirjojen dialogien tarkastelua. – Johanna Tanner, Marja Kokkonen (Toim.). *Suomenopetus, kielitaito ja tutkimus. Kakkoskieli 6*. Helsinki: Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.
- Tanner, Johanna 2012. Rakenne, tilanne ja kohteliaisuus: Pyynnöt S2-oppikirjoissa ja autenttisissa keskusteluissa. Helsinki: Helsingin yliopisto.
- Zimmerman, Cheryl Boyd, Norbert Schmitt 2005. Lexical questions to guide the teaching and learning of words. – *The CATESOL Journal* 17 (1), 1–7.

**Tuija Määttä**

Umeå universitet, Institutionen för språkstudier

S-901 87 Umeå, Sweden

[tuija.maatta@finska.umu.se](mailto:tuija.maatta@finska.umu.se)



# The influence of vocabulary in teaching material on texts written by Swedish-speaking beginners-level learners of Finnish

TUIJA MÄÄTTÄ

Umeå University

This paper presents a study that focuses on how teaching materials used in teaching beginners-level learners of Finnish might influence the learning process in relation to writing skills.

The present research questions are: 1) Are there any signs that the vocabulary in the texts written by Swedish-speaking students has been influenced by the teaching material? 2) What kinds of influence can be seen in the student's texts? 3) How do the students use the different word classes? and 4) Which are the most frequent words in the student's texts?

The research is based on two corpora:

- 1) 17 beginners-level Swedish-speaking students wrote three texts each:
  - Text A: *Write an e-mail to your friend*, 1,722 tokens in total (based on chapter 3 in the textbook)
  - Text B: *Describe a fantasy or real person*, 2,599 tokens in total (based on chapter 4 in the textbook)
  - Text C: *Describe one ordinary day in your life*, 1,929 tokens in total (based on chapter 5 in the textbook).

This electronic corpus, having generated a total of 6,250 tokens, is a part of the Swedish subcorpus of the International Corpus of Learner Finnish, ICLFI, which has been compiled since 2007 as part of the *Corpus study on language-specific and universal features in learner language* project.

2) Vocabulary lists for the first five chapters of the textbook *Suomen mestari 1* ("Mastery of Finnish 1") were compiled at the Umeå University Department of Language Studies, Finnish language. The vocabulary lists contain all of the words which appear in the texts, dialogues, etc. of each of the five chapters.

The words and phrases are provided in exactly the same form as they appear in the chapters.

The electronic corpus is investigated using comparative and quantitative methods. The comparison and analysis of the vocabulary in the students' texts and the vocabulary in the textbook is enabled through the use of the computer programmes WordList, which is included in the WordSmith Tools software package.

The results show that those students who read the chapters in the textbook and then produced a text did use words, phrases and even whole sentences from the chapters. In total, 34% of the occurrences of such words were drawn from the textbook. The comparison of word classes in the students' texts with the vocabulary in the textbook chapters shows that nouns and verbs are the most frequent word classes in both corpora. The most common words in the students' texts are the conjunction *ja*, meaning 'and', and the verb *olla*, meaning 'to be'. The pronouns *minä*, 'I', and *hän*, 'he, she', occur quite frequently, as well.

The results of this study show that the texts the Swedish-speaking students have produced were influenced by the textbook to some degree. This investigation focuses on vocabulary only, and the students' texts contain a number of words written in basic or inflected forms. These forms can also be found in the textbook. The possible interpretations of this which are considered here include: 1) the students have genuinely learned the words and word forms and know how to use them correctly; 2) the students have simply copied the words and word forms directly from the textbook; or, 3) the students received assistance from someone who knows the words and word forms and how to use them.

On the basis of just this study, however, it is difficult to draw any definite conclusions. A further study with a larger corpus and with computer programmes better suited to conducting such analyses would be needed in order to arrive at a deeper understanding.

**Keywords:** corpus-based research; vocabulary; learner Finnish; word classes; word frequency; Finnish; Swedish