

Suomen kielen *U*-verbit taitotason indikaattorina

KIRSTI SIITONEN

Turun yliopisto

MAISA MARTIN

Jyväskylän yliopisto

Tiivistelmä. Artikkelin lähtökohtana on kysymys lingvistisistä indikaattoreista eri funktionaalisia kielitaitotasoa osoittamassa. Tutkimuskohteena ovat suomen kielen *U*-verbit, joiden on aieman tutkimuksen perusteella todettu tuottavan ongelmia hyvin edistyneillekin suomenoppijoille. *U*-verbit ovat intransitiivisia verbijohdoksia, joiden kantaverbeillä on tavallisesti *A*-vartalo, esim. *vaihtaa* > *vaihtua* ja *jatkaa* > *jatkaa*. *U*-verbit voivat olla esimerkiksi refleksiivisiä tai automatiivisia. Automatiivisuudella tarkoitetaan, että jotakin tapahtuu ei-intentionaalisesti tai ilman elollista agenttia.

Aineistona ovat otokset kolmesta suomenoppijoiden korpuksesta: 12–16-vuotiaiden koululaisten kirjoitelmista (EVK:n tasoilta A1–B2), Yleisten kielitutkintojen testeistä (tasot A1–C2) ja maisteriopiskelijoiden akateemisista kirjoitelmista (tasot B2–C2). *A*- ja *U*-verbiparit (42 paria) on kerätty yhteensä n. 182 000 saneen aineistosta.

Tutkimustulosten mukaan *U*-verbit näyttävät olevan kielitaidon tason indikaattoreita kahdella tavalla: *U*-verbien variaatio rikastuu ja esiintymämäärät kasvavat taitotasolta toiselle, erityisesti akateemisissa tekstilajeissa. Toisaalta myös virheiden määrä niiden käytössä lisääntyy, mikä johtaa epätavalliseen tilanteeseen, jossa virheet eikä niiden puuttuminen on merkki kielitaidon korkeasta tasosta. Alemmilla kielitaidon tasoilla käytössä on pääasiassa korrektisti käytettyjä valmiita kiteymiä.

Avainsanat: ilmaisun tarkkuus; kielitaidon tasot; kielitaidon tason indikaattori; *U*-verbit

1. Johdanto

Uuden kielen oppiminen sisältää aina sekä kielellisen aineksen että sen käytön omaksumista. Sanoja ja rakenteita on painettava muistiin joko erillisinä yksiköinä ja abstrakteina sääntöinä tai erilaisina konstruktioina, joihin sisältyy sanastollista, rakenteellista ja semanttista tietoa. Oppimisprosessi on yksilöllinen, dynaaminen ja varioiva eikä etene lineaarisesti jatkuvana kasvuna vaan sisältää nousuja ja laskuja. Toisen kielen oppimisen tutkimus pyrkii löytämään näiden yksilöllisten prosessien yhteisiä piirteitä. Vaikka kenenkään kielenoppimisprosessi ei joka yksityiskohdaltaan noudata tällaista keskimääräistä polkua, sen tutkimisella on ansionsa, sillä kielenopetus, oppimateriaalit ja kielitaidon arviointi kohdistetaan lähes aina ryhmille. Tässä artikkelissa tarkastellaankin oppijoiden kirjallisia suorituksia siltä kannalta, mitä niissä tyypillisesti esiintyy tai ei esiinny tietyllä taitotasolla, kun tutkittavana on suuri joukko kirjoitelmia. Taitotasoa voi osoittaa myös se, millaisena jokin ilmiö esiintyy.

Jotakin kielen ilmiötä – tässä *U*-verbiin käyttöä – tarkastellaan toisen kielen oppimisen tutkimuksessa yleensä suhteessa kyseisen kielen hallinnan yleiseen tasoon. Tavallisia määrittelyjä ovat esimerkiksi alkeistaso, keskitaso ja edistynyt taso. Useimmiten nämä on määritelty saadun opetuksen tunti- tai vuosimäärän pohjalta, harvoin jollakin yleistä kielitaitoa mittaavalla testillä. Jos taustalla on testi, se koskee tavallisesti yksittäisen oppijan yleistä kielitaitoa. Tämän artikkelin aineisto sen sijaan on arvioitu tiettyä kielitaidon tasoa edustavaksi joko kunkin yksittäisen kirjoitelman osalta tai kirjoittajan yleisen taitotason perusteella (ks. tarkemmin luku 5). Arvioinnin lähtökohtana on Kielten oppimisen, opettamisen ja arvioinnin yhteisen eurooppalaisen viitekehyksen (EVK 2004) tarjoama taitotasoajattelu, josta on valittu kirjoitetun kielen käyttöä erittelevät kuvaimet (ks. tarkemmin Alanen ym. 2010 sekä tämän

artikkelin luku 5). Tutkitut kirjoitelmat on arvioitu näin muodostetulla funktionaalisella asteikolla eli sen mukaan, mitä kirjoittaja tekstin perusteella osaa tai ei osaa kyseisellä kielellä kirjoittamalla tehdä.

Toisen ja vieraan kielen oppimisen tavoitteet on suomalaista kielenopetusta ohjaavissa opetussuunnitelmien perusteissa (OPS, Opetushallitus 2004) ja Yleisten kielitutkintojen tavoitteissa (YKI) asetettu funktionaalisesti. Kielellä toiminen vaatii kuitenkin aina myös lingvistisiä resursseja, joiden tarjoaminen opetuksessa perustuu tällä hetkellä pitkälti käytännön kokemukseen ja oletuksiin, ei tutkimustuloksiin luontaisesta tai opetuksen avulla saavutetusta oppimisjärjestyksestä. Tutkimalla sitä, mikä eri funktionaalisille taitotasoille arvioiduille kirjoitussuorituksille on kielellisesti ominaista, voidaan saada tietoa siitä, missä järjestyksessä kielellisiä resursseja todellisuudessa otetaan haltuun. Tässä artikkelissa käsitellyt *U*-verbit ovat eräs tällainen kielellinen resurssi, joka saattaa olla yhteydessä kielen käytön edistymiseen ja saattaisi siis olla kirjallisen kielen käyttötaidon kehitystä ja kielen sanaston ja rakenteiden oppimista yhdistävä seikka, taitotason indikaattori.

U-verbit (esim. *jatkua*, *vaihtua*) käyttäytyvät semanttisesti ja syntaktisesti eri tavalla kuin kantaverbinsä (esim. *jatkaa*, *vaihtaa*), mutta morfofonologinen ero on melko huomaamaton. Niiden käyttö ja tulkinta on osoittautunut edistyneillekin S2-kirjoittajille hankalaksi (Siitonen 1999). Alkuvaiheen oppijoiden kielestä niitä ei ole tutkittu. Eroja oppimisessa saattaa aiheuttaa myös oppimisympäristö. Suomea ulkomaisessa yliopistossa opiskellut saattaa tuntea *U*-verbit teoriassa, mutta ei osaa kirjoittaessaan ja sisältöön fokusoidessaan soveltaa tietoaan käytäntöön. Suomalaisessa ympäristössä elävien nuorten toisen kielen oppijoiden voisi taas kuvitella oppivan ne samalla tavalla kuin muutkin verbit, todennäköisesti omaksumalla käyttöönsä joitakin valmiita kokonaisuuksia, joiden pohjalta konstruktiot abstrahoituvat ja erot *jatkaa*- ja *jatkua*-tyyppisten verbien välillä vähitellen selkiytyvät, jos vain sanahahmot erottuvat riittävän tarkasti.

Esittelemme seuraavassa ensin indikaattorin käsitettä, sitten määrittelemme, mitä *U*-verbeillä tässä tarkoitetaan. Sen jälkeen nostamme

esiin syitä, jotka saattaisivat vaikuttaa siihen, että *U*-verbien käyttö voisi olla taitotasoa indikoiva muuttuja. Luvussa 5 esittelemme aineiston ja sen jälkeen tutkimuksemme tuloksia kahdesta näkökulmasta: *U*-verbien käytön taajuus ja tarkkuus sekä niiden käytön laadullinen muutos taitotasolta toiselle. Lopuksi pohdimme, onko *U*-verbien käytön määrässä, virheettömyydessä tai käyttötavoissa sellaisia taitotasoittain vaihtelevia piirteitä, että niitä voitaisiin pitää taitotason indikaattoreina.

2. Taitotason indikaattorit

Kielenoppimisen edistymistä voi tarkkailla monin tavoin. Funktionaaliset taitotasokuvaukset ovat yksi tällainen tapa. Lingvistikseen näkökulmasta taas keskustellaan paljon siitä, kuinka monta kertaa jonkin ilmiön on esiinnyttävä oppijan kielessä, että hänen voitaisiin katsoa oppineensa (ks. esim. Pallotti 2007). Myös kielellisten ilmausten tarkkuutta eli virheettömyyttä voidaan käyttää oppimisen mittarina. Joskus vaaditaan täydellistä kohdekielenomaisuutta, toisinaan asetetaan rajaksi vaikkapa 80 %:n tarkkuus pitäen mielessä se, että ensikielisillekin kirjoittajille sattuu lipsahduksia. Myös 50 %:n tarkkuutta on käytetty, koska se kertoo kuitenkin jo kyseisen ilmiön melko hyvästä hallinnasta (esim. Glahn ym. 2001). Toisaalta yhtäkin esiintymää voidaan pitää merkinä oppimisesta (esim. Pienemann 1998), jos voidaan näyttää, ettei kyse ole ulkoa opitusta fraasista. Konstruktionäkökulmasta (ks. esim. Goldberg 1995, Tomasello 2003) kielenoppimista tarkastelevat taas lähtevät siitä, että jonkin konstruktion esiintyminen ensimmäisen kerran osoittaa jo sen oppimista (esim. Martin ym. 2010) eikä siten ole järkevää asettaa oppimisen rajaa esiintymiskertojen perusteella.

Näissä oppimisen määrittelyissä on kyse aina yksilön oppimisesta. Tässä artikkelissa taas tarkastellaan suurta joukkoa oppijoiden tekstejä, ei yhden oppijan kehitystä. Tästä syystä käytämme nimitystä *taitotason indikaattori*. Sillä viitataan kielelliseen piirteeseen (johonkin rakenteeseen tai sanastolliseen ilmiöön tai vaikkapa pragmaattisen konvention käyttöön), jonka esiintyvyys tietyllä funktionaalisella taitotasolla poik-

keaa sen esiintyvyydestä jommallakummalla viereisellä taitotasolla. Esimerkiksi transitiivirakenteiden määrä tuhatta sanetta kohden on EVK:n tasolla B1 (ks. taitotasojen esittely liitteessä 1) selvästi suurempi kuin edellisellä taitotasolla A2 (Martin ym. 2010). Tällöin voi sanoa, että transitiivirakenteiden runsastuminen on tason B1 indikaattori. Vastaavasti transitiivirakenteissa esiintyy selvästi vähemmän virheitä seuraavaksi ylemmällä tasolla B2 kuin tasolla B1. Tällöin transitiivirakenteiden suhteellinen virheettömyys on taitotason B2 indikaattori. Joskus sama ilmiö voi olla indikaattori sekä alemman että ylemmän tason suhteen. Esimerkiksi imperfektin virheelliset kieltorakenteet kumuloituvat tasolle A2 (Martin 2008). Niitä on huomattavasti vähemmän sekä tasolla A1, jolla oppijat harvoin edes yrittävät käyttää kieltorakennetta imperfektissä, että tasolla B1, jolla taas kierto osataan useimmiten muodostaa myös imperfektissä. Taitotason indikaattorina voi siis olla yhtä lailla jonkin ilmiön esiintyminen kuin sen puutekin. Samoin niin virheiden esiintyminen kuin niiden puutekin voivat indikoida taitotasoa. Myös kolmas tässäkin artikkelissa tarkasteltava tekijä, ilmiön käytön monipuolistuminen, voi toimia indikaattorina, vaikka se harvoin on määrällisesti mitattavissa. Sen luonne indikaattorina on pikemmin laadullinen; tietynlaisia kompleksisia rakenteita alkaa esiintyä vasta ylemmillä tasoilla (ks. Reiman 2011).

Indikaattorit ovat tuskin koskaan absoluuttisia. Yksittäisenä tapauksena miltei mitä tahansa saattaa esiintyä millä tasolla tahansa, jos aineisto on riittävän suuri. Toisinaan kuitenkin yksittäinenkin ilmaus voi näyttäytyä indikaattorina. Esimerkiksi tämän tutkimuksen keskisuurta Cefling-aineistosta (ks. luku 5) ei löydy yhtään *vaikka*-konjunktiota ennen taitotasoa B1 (Martin 2010). Näin ollen tämän konjunktion voi sanoa olevan melko vahva merkki taitotason B saavuttamisesta. Tavallisesti kuitenkin kyse on taitotasojen välisestä erosta, joka poikkeaa lineaarisesta kehityksestä taitotasolta toiselle. Onhan selvää, että minkä tahansa ilmiön esiintyvyys lisääntyy, tarkkuus paranee ja kompleksisuus kehittyy, kun siirrytään taitotasolta A1 tasolle C2. Jotta muutosta voisi pitää merkittävänä taitotason indikaattorina, vierekkäisten taitotasojen

eron on kuitenkin oltava huomattava, määrällisten erojen osalta mieluiten tilastollisesti merkitsevä. Harvoin esiintyvien ilmiöiden osalta tähän ei kuitenkaan aina voida päästä. Silti muutos ilmiön esiintyvyydessä tai tarkkuudessa voi olla merkittävä; samoin laadullinen muutos voi olla havaittava siitä huolimatta, ettei sitä voida määrällisesti mitata.

Mikään kielellinen ilmiö yksinään ei tietysti todista kirjoitelman olevan tietyllä funktionaalisella tasolla. Edes tulevaisuudessa, kun mahdollisia kielen piirteitä on nykyistä kattavammin tutkittu, kaikkien tunnettujen indikaattorien yhtäaikainenkaan läsnäolo ei varmista kirjoitelman olevan millään tietyllä funktionaalisella tasolla – kysehän on nimenomaan kahdesta eri kehityslinjasta, lingvistiksestä ja funktionaalisesta. Yhteyksien etsiminen näiden kahden kehityslinjan välillä on silti tarpeellista, sillä se auttaa oppijoita ja opettajia suuntaamaan työskentelyään: kun tavoitteet on asetettu funktionaalisesti, kielenopetuksessa on järkevää keskittyä sellaisiin sanastollisiin ja rakenteellisiin seikkoihin, joiden tiedetään olevan tarpeen seuraavalle tasolle pääsemiseksi. Kaikki tason B1 kirjoittajat eivät koskaan käytä samoja kielen resursseja, mutta jos suuri osa heistä käyttää joitain niistä, näiden sisällyttäminen opetusohjelmaan auttaa oppijoita pääsemään tasolle B1.

Joistakin suurista kielistä on olemassa runsaasti toisen kielen oppimisen kulun tutkimusta. Sen perusteella on mahdollista ennustaa, mitkä ilmiöt voisivat toimia taitotason indikaattoreina. Esimerkiksi englannin oppimisen osalta indikaattoreita ovat tämän tutkimuksen tapaan, vaikka hieman eri termein, määritelleet Hawkins ja Buttery (2010) lähtökohtanaan aiempi englannin oppimisen tutkimus. Suomen kielen oppimisjärjestyksestä ei ole samanlaista systemaattista ja tutkimusperäistä tietoa, jonka pohjalta voitaisiin valita mahdollisia indikaattoreita. Kokeneilla opettajilla ja kielitaidon arvioijilla on toki paljon kokemusperäistä tietoa siitä, millaisia kielellisiä ilmiöitä eri tasoilla esiintyy. Kokemusperäistä tietoa on käytetty hyväksi myös Cefling-hankkeen tutkimuskohteita valittaessa. Ennalta ei kuitenkaan voi tietää, mikä kielen ilmiö toimii indikaattorina, mikä ei. Useimmat tähän asti tutkitut potentiaaliset indikaattorit on valittu niin, että ne erottelisivat alimpien tasojen

(A1–B1) välillä, sillä suuri osa oppijoista on näillä tasoilla. Sitä mukaa kuin suomenoppijoiden määrä ja koulutustaso nousee, lisääntyy myös tarve ylimpiä tasoja erottelevien indikaattoreiden löytämiseen. Tähän tutkimukseen valittiin *U*-verbit, koska tiedämme aiemman tutkimuksen pohjalta (Siitonen 1999), että ne tuottavat vaikeuksia pitkällekin edistyneille oppijoille.

3. *U*-verbit

U-verbit on suomen kielen johdosryhmä, jota on tutkittu melko paljon. Niiden merkitys riippuu kulloisestakin käyttötilanteesta ja voi olla esimerkiksi passiivinen, refleksiivinen tai automatiivinen. Tiukan rajauksen mukaan *U*-verbejä ovat vain deverbaaliset *U*-verbijohdokset, joilla parina on *A*-verbi: *peittyä* ← *peittää*, *jatkoa* ← *jatkaa*, *siirtyä* ← *siirtää*. Toisinaan *U*-verbillä ei ole suoraa suhdetta *A*-verbiin, vaan kantasanaan voi esimerkiksi olla *e*-vartaloinen verbi: *särkyä* ← *särkeä*. Löyhemmän rajauksen mukaan *U*-verbeinä voidaan pitää myös *UtU*-johdoksia. Toisinaan, kuitenkin melko harvoin, sananjohdon avulla muodostuu selväpiirteinen sarja *A*-verbistä *U*-verbin kautta *UtU*-johdokseen: *pelastaa* → *pelastua* → *pelastautua*. (Kulonen-Korhonen 1985; Koivisto 1991: 65, 91–93; 1995: 37–49; Siitonen 1999: 84–94.)

Olennaista *A*- ja *U*-verbien suhteessa on transitiivisuus–intransitiivisuus-dikotomia, jossa *A*-verbit ovat transitiivisia ja saavat objektin. Näin ollen vastaavia ilmiöitä edustavat monenlaiset leksikaaliset transitiivi–intransitiivi-parit kuten *katkaista* vs. *katketa* tai *pudottaa* vs. *pudota*. Useiden *U*-verbien käytölle on ominaista, että niiden johtamisessa ja taivutuksessa on rajoituksia: esimerkiksi *jA*-johdokselle *vaihtuja* ja taivutuspassiiville *vaihdutaan* ei löydy helposti luontevaa käyttöä. Oppijoille *U*-verbien johtosuhteiden aukkoisuus ja käytön rajoitukset sekä merkityksen läheisyys *A*-verbien kanssa tuottavat ongelmia. Ensikieliset suomenpuhujat hallitsevat *U*-verbit mutta eivät välttämättä käsitä verbi-parien jäseniä eri verbeiksi. Lisäksi äänteellisesti lähekkäiset verbit aiheuttavat helposti ensikielisilläkin lipsahduksia. (Siitonen 1999: 237, 242.)

4. *U*-verbit opittavana ilmiönä

Luvussa 3 on esitetty karkeasti intransitiivisten *U*-verbien suhde niiden transitiivisiin pareihin. Tällaisenaan suhde vaikuttaa helposti omaksuttavalta ja ymmärrettävältä. Aivan prototyyppisissäkin tapauksissa oppijaa saattavat kuitenkin hämmentää niin rektio- kuin agenttisuhteetkin.

- (a) Olen varma, että siihen *sisältää* myös kaipausta.
- (b) Hän on tehnyt oppikirjan, joka *sisältyy* paljon tekstejä
- (c) --- hän (Minna Canth) jatkoi isänsä *perustumaa* kauppaliikettä
- (d) Ovi *avasi* itsestään.
- (e) Minulla oli kerran englantilainen sateenvarjo, se *avasi* nopeasti ja oli vaarallinen.

(Siitonen 1999: 45–49.)

Näiden melko selvien tapausten (a–e) lisäksi kielessä on paljon ei-prototyyppisiä tapauksia. Osa suomen kielen verbeistä on luontaisesti hankalia transitiivisuuden ja intransitiivisuuden selvän työnjaon kannalta. Erityisen paljon ongelmia tuottaa verbi *muuttaa* (f) ilmaistessaan asuinpaikan vaihtoa ja käyttäytyessään intransitiivisesti siitä huolimatta, että se on *A*-verbi. (Siitonen 1993.) Verbi on jo opintojen alkuvaiheessa hyvin frekventti. Silloin se ehkä vielä hallitaan, mutta *A–U*-dikotomian tultua tutuksi käyttö hämärtyy.

- (f) Perheeni *muuttui* Vaasaan.
 - (g) Minun pää *särkyi*.
- (Siitonen 1999: 139.)

Esimerkissä (g) hämmentää lisäksi tunnekausatiivi-ilmausten kaltainen *päästä särkee* -rakenne.

Edellä luvussa 3 on esitelty prototyyppisiä tapauksia. Todellisessa kielenkäytössä esiintyy kuitenkin paljon epätäydellisiä sarjoja (h, i). Verbillä saattaa olla useita semanttisia tulkintoja eikä kaikille löydy vastinparia.

- | | | |
|--------------------|-------------------|----------------|
| (h) päättää kokous | kokous päätetään | kokous päättyy |
| päättää asiasta | asiasta päätetään | – |

(i) painaa kankea	kankea painetaan	kanki painuu
painaa lehteä	lehteä painetaan	–
painaa 50 kg	–	–

(Vrt. Siitonen 1999: 294.)

Oppijan ensikielen siirtovaikutus on myös otettava huomioon. Usein omasta kielestä puuttuu suomessa esiintyvän dikotomian malli. Seuraavassa (j) on viisi hyvin frekventtiä *A*-verbiä, joilla suomessa on *U*-johtiminen pari. Kummankin vastine on verbin jälkeen mainituissa kielissä täsmälleen sama.

(j) <i>jatkaa</i>	englanti, hollanti, tanska, norja, ruotsi, espanja, italia, bulgaria
<i>vaihtaa</i>	englanti, saksa, norja, espanja
<i>muuttaa</i>	englanti, hollanti, norja, espanja, bulgaria
<i>siirtää</i>	englanti, norja
<i>sytyttää</i>	englanti, ruotsi

(Siitonen 1999: 387–389.)

Lisäksi indoeurooppalaisissa kielissä sekä suomen kielen *U*-verbejä että suomen kielen taivutuspassiivia voi vastata sama ilmaus. Suomeksi käännettäessä taas ei voi ajatella, että indoeurooppalainen passiivi-ilmaus käännettäisiin kummalla tavalla tahansa. Suomen passiivin oletetaan ensisijaisesti sisältävän henkilötekijän ja intentionaalisuuden: *kuppi särjettiin* vs. *kuppi särkyi* vrt. engl. *the cup was broken*.

Kun vielä loogisesti transitiivisuuteen yleensä liitetään intentionaalisuus (Hopper–Thompson 1980: 279–280, 287–288, 294), ei-intentionaalisesti käytetyt transitiiviverbit ovat omiaan aiheuttamaan sekaannuksia: *Hän katkaisi jalkansa* (Siitonen 2002).

Vaikeinta oppijalle kuitenkin tuntuu olevan ns. automatiivisuus, esimerkiksi *lapaset vaihtuivat päiväkodissa*. Tekijätöntä ilmausta on vaikea hahmottaa. Teon takana lienee ihminen, mutta hän ei kuitenkaan ole tehnyt mitään tietoista valintaa. Useiden kielten puhujat pyrkivät etsimään lauseesta intentionaalista agenttia tai kirjoittamaan lauseeseen sellaisen. Pyrkimystä liittää agentin rooli eksplisiittiseen lauseenjäseneen on pidetty oppijoille tyypillisenä kielestä riippumatta. (Ks. Ellis 2008: 474–476.)

Automatiivisten ilmausten merkityskomponenttina on myös tahattomuus, joka kieliopillisena kategoriana on monelle aivan uusi. Tahattomuutta ilmaisevan verbijohdoksen käyttö (*lapaset vaihtuivat*) passiivin sijaan (*lapaset vaihdettiin*) on kohtelias valinta. Tässä *U*-verbin käyttö ei kuulosta eksplisiittiseltä kohteliaisuudelta, mutta pragmaattiselta kannalta tällainen valinta osoittaa hienotunteisuuden hallintaa. (Ks. Martin 2000.)

Tässä käsitellyt verbit taivutuksineen sisältävät usein liian paljon vaikeasti havaittavia eroja (k). Suomen kielessä on monenlaisia vokaalivaihteluita, joihin *U*-johtimet helposti sekaantuvat. Toisen kielen puhuja ei välttämättä havainnoi, milloin loppuvokaalina on *o*, milloin *u*. Hän ehkä kiinnittää huomionsa vain sanan alkuun. (Ks. Siitonen 1999: 311.) Myös taivutusaineisiin sisältyvät vokaalit saattavat hämmentää.

(k) <i>A</i> -verbejä		<i>U</i> -verbejä	
jatkaa	jatkoi	jatkuu, jatkui	
siirrettyä	siirrettävä	siirryttävä	(Ks. Siitonen 1999: 350.)

Vokaalien sekaantumista osoittanevat seuraavatkin esimerkit (l, m, n). Niissä on tuskin kyse tietoisesta sananvalinnasta.

- (l) Puhelinsoiton jälkeen hän *jatkuu* suutumista.
 (m) Voin *vaihtua* paikkaa.
 (n) En *löydy* autoa. (Siitonen 1999: 138–139.)

A- ja *U*-verbien johtosuhteiden epäsäännöllisyys, niiden ilmaisemien erojen, varsinkin tahattomuuden, puuttuminen oppijan ensikielestä sekä kyseisten verbien fonologinen samankaltaisuus aiheuttavat siis oppijoille monenlaisia ongelmia.

5. Aineistot

Minkä tahansa yksittäisen sanan esiintyminen jossakin tekstissä on aina sattumanvaraista. Kirjoittajallahan on yleensä lukuisia mahdollisuuksia ilmaista ajatuksensa, vaikka kirjoituksen aihe ja mahdollinen tehtävänanto tietenkin vaikuttavat sanojen valintaan. Tästä syystä pyrimme hakemaan *U*-verbejä ja niiden *A*:llisia pareja tätä artikkelia varten mahdollisimman suuresta oppijankielen kirjoitelma-aineistosta. Käytettävissämme on kolme korpusta, jotka edustavat 12–16-vuotiaiden yläkoululaisten, Yleisiä kielitutkintoja suorittavien aikuisten ja Turun yliopistossa opiskelevien suomen ja sen sukukielten maisteriohjelman opiskelijoiden tekstintuottamista. Seuraavassa kukin aineisto on esitelty, minkä jälkeen on aineistomäärät sisältävä taulukko (taulukot 1–3).

Koululaisaineisto on kerätty Cefling-hanketta¹ (2007–2009) varten suomea toisena kielenä vuosiluokilla 7–9 opiskelevilta maahanmuuttajataustaisilta oppilailta eri puolilta Suomea. Oppilaiden ensikieliä on 35, joten osallistujat eivät edusta mitään tiettyä kielellistä oppijaryhmää. Kirjoitelmat edustavat neljää tekstilajia: epämuodollinen viesti (kaverin kanssa sovitun tapaamisen peruuttaminen, kotitehtävien kysyminen opettajalta), muodollinen viesti (viesti verkkokauppaan huonosti toimivasta tietokonepelistä), mielipide (kännyköiden käyttö koulussa tai vanhempien lupa valvoa lastensa internetin käyttöä) ja kertomus (pelottava tai hauska tapahtuma). Yhdeltä kirjoittajalta on mukana korkeintaan neljä tekstiä, yksi kutakin tekstilajia.

Aineiston kirjoitelmat on arvioitu mahdollisimman luotettavasti Eurooppalaisen kielenoppimisen, -opetuksen ja kielitaidon arvioinnin viitekehykseen perustuvalla EVK-asteikolla (liite 1) taitotasolle A1–B2. Tähän tutkimukseen on otettu mukaan vain ne kirjoitelmat, joiden taitotasosta arvioijat olivat (lähes) yksimielisiä (arviointiprosessista tarkemmin ks. Alanen ym. 2010). Arviointiasteikkoon oli valittu vain

¹ Cefling-hanke eli *Linguistic Basis of the Common European Framework for L2 English and L2 Finnish* oli Suomen Akatemian ja Jyväskylän yliopiston rahoittama tutkimushanke, ks. verkkosivu <https://www.jyu.fi/hum/laitokset/kiellet/cefling>.

funktionaalaisia kirjoittamisen kuvaimia, jotta vältettäisiin mahdollisimman tarkoin toiminnallisen ja lingvistisen arvoinnin päällekkäisyys. C-tason kirjoitelmia ei koululaisilla ollut riittävästi, jotta niistä olisi voitu muodostaa oma luokkansa. Myöskään B2-tason kirjoitelmia ei ollut riittävästi, jotta niiden määrä olisi yhdenmukainen muiden taitosojen kanssa. Tämä ei tietenkään ole yllättävää, onhan suomi toisena kielenä -oppimäärän tavoitteet peruskoulun päättyessä asetettu tasolle B1 (OPS, Peruskoulun opetussuunnitelman perusteet 2004). Kouluaineisto on arvioitu myös Peruskoulun opetussuunnitelman perusteiden liitteenä olevalla kymmenportaisella asteikolla, mutta näitä arviointeja ei tässä tutkimuksessa ole käytetty.

TAULUKKO 1. *Kouluaineiston sana- ja tekstimäärät taitotasoittain*

Taso	Sanoja	Tekstejä	Tekstien keskipituus sanoina
A1	3 479	115	30,3
A2	9 044	199	45,4
B1	11 317	184	61,5
B2	2 223	29	76,7
Yhteensä	26 063	527	49,5

Cefling-hankkeen yhtenä tavoitteena oli myös verrata aikuisten ja koululaisten S2-kirjoitustaitoja. Tästä syystä kouluaineistoon valittiin sellaisia tekstilajeja, joiden hallintaa arvioidaan myös Yleisten kielitutkintojen (YKI) kokeissa ja jotka soveltuvat myös koululaisten kirjoitettaviksi. Näitä ovat muodolliset ja epämuodolliset viestit sekä mielipidekirjoitukset. Aikuisilta on mukana korkeintaan kolme saman kirjoittajan tekstiä. YKI-aineistoa on vuosien mittaan kertynyt tuhansia suorituksia, mutta tähän tutkimukseen valittiin vuosien 2007 ja 2008 suorituksista koululaisaineiston suurimpia yksimielisesti arvioituja ryhmiä vastaavat määrät eri taitotasoilta. Kertomuksia ei YKI-aineistossa ole, mutta koululaisilta kerättiin myös kertomuksia, sillä etenkin 7.-luokkalaisille se on tutuin tekstilaji; esimerkiksi mielipidekirjoitusta harjoitellaan yleensä

vasta 8. luokalla. YKI-aineiston ovat arvioineet koulutetut YKI-arvioijat. Myös valtaosa koululaisaineiston arvioijista oli saanut YKI-koulutuksen. Lisäksi arvioijia koulutettiin Cefling-hankkeen piirissä. YKI:n suorittajat edustavat erittäin monia eri ensikieliä ja heidän koulutustasonsa vaihtelee suuresti, mutta kaikki ovat aikuisia. Kummankin aineiston kirjoitelmat on kirjoitettu valvonnan alaisena ilman apuneuvoja, joko koululuokassa tai YKI-kokeissa.

TAULUKKO 2. YKI-aineiston sana- ja tekstimäärät taitotasoittain

Taso	Sanoja	Tekstejä	Tekstien keskipituus sanoina
A1	4 974	113	44,0
A2	5 702	103	55,4
B1	10 861	126	86,2
B2	9 080	108	84,1
C1	11 550	117	98,7
C2	10 852	102	106,4
Yhteensä	53 019	669	79,3

Maisteriopiskelijoiden kirjoitelmia kerätään Turun yliopistossa Edistyneiden suomenoppijoiden korpus -hankkeeseen (LAS2). Kolmantena aineistona tässä artikkelissa on otos kyseisen LAS2-hankkeen aineistosta. Aineisto sisältää tenttivastauksia, esseitä ja opinnäytetöiden ensimmäisiä versioita. Tenttivastaukset ovat suunnilleen muiden tässä tutkittujen aineistojen tekstien laajuisia; esseet ja tutkielmaluonnokset ovat huomattavasti pitempiä. LAS2-aineistoa ei ole arvioitu taitotasoittain, mutta se edustaa tasoja B2–C2 sekä akateemista suomea edustavien tekstilajien että ohjelman sisäänottokriteerien perusteella. Pääosa informanteista on aloittanut suomen kielen opinnot ulkomaisessa yliopistossa. Tässä käytetyssä aineistossa on 9 eri kirjoittajaa (tenttivastausaineistossa 9, esseaineistossa 6 ja tutkielmaluonnosaineistossa 4); heidän ensikielensä ovat viro, komi, ruotsi, englanti, tšekki, venäjä (2) ja liettua (2). Kultakin kirjoittajalta on useampia tekstejä tässä tarkastellussa aineistossa. Vain

aineiston tenttivastaukset on kirjoitettu valvonnan alaisena, ja tällöinkin on useissa tapauksissa ollut mahdollista käyttää sanakirjaa. Esseet ja opinnäytetöiden ensimmäiset versiot on kirjoitettu itsenäisesti valvomattomissa olosuhteissa. Opinnäytetöiden ensimmäisiä versioita on pro gradu -tutkielman useista eri luvuista, joita on kirjoitettu eri aikoina. Ensimmäinen aineistoon tallennettu osa opinnäytetyöstä on tutkielmaseminaarin työ, jonka aineksia – usein jo idiomaattisemmassa asussa – on myös myöhemmin kirjoitetuissa luvuissa.

TAULUKKO 3. LAS2-aineiston sana- ja tekstimäärät

Tekstilaji	Sanoja	Tekstejä	Tekstien keskipituus sanoina
Tenttivastaukset	32 903	234	140,6
Esseet	19 470	22	885,0
Tutkielmaluonnokset	50 770	15	3 384,7
Yhteensä	103 143	271	380,6

Jäljempänä näistä korpuksista lainattujen esimerkkien lähteet on merkitty koodeilla, joita kyseisissä hankkeissa on käytetty. Niiden perusteella on mahdollista tarvittaessa jäljittää teksti ja kirjoittajan taustatiedot.

Valitaksemme korpuksista haettavat verbit kävimme ensin läpi korpuksia ja keräsimme joukon verbejä, joita niissä esiintyi. Ne ja niiden parit (esim. *aiheuttaa/aiheutua*) kokosimme listaksi, joka on liitteenä 2. Kaikki parit eivät ole johdossuhteiltaan yhtenäistä *jatkaa/jatkua*-tyyppiä (ks. luku 3). Jätimme kuitenkin pois tavallisetkin *U*-verbit, ellei niillä ole läpinäkyvää johtosuhdetta toiseen samankantaiseen verbiin (esim. *tapahtua, huolestua, käyttäytyä*). Sen jälkeen tarkistimme valitsemiemme verbien esiintymät systemaattisesti korpuksissa ja taulukoimme esiintymät aineistoittain ja taitotasoitain. Kävi ilmi, että joillakin verbeillä on runsaasti fraasinomaista käyttöä, esimerkiksi *nähdä*-verbin esiintymistä suuri osa etenkin alemmilla tasoilla on viestin lopussa olevia *nähdään* (*huomenna/koulussa*) -tyyppisiä. Nämä on poistettu *nähdä*-verbin

kokonaismääristä, joissa on siis mukana vain ne tapaukset, joissa *nähdä* on kokonaisen lauseen predikaattiverbinä. Vastaavasti *kuulua*-verbi on usein osana fraasia *mitä kuuluu*. Myös nämä on jätetty pois laskuista.

Aineisto ei tietenkään kata likikään kaikkia suomen *A*- ja *U*-verbi-pareja, ei edes tavallisimpia, vaan siihen on valikoitunut verbejä aineistossa esiintyvien tekstilajien ja tekstien sisällön perusteella. Tulokset eivät siis kerro kaikkien *A*- ja *U*-verbien käytöstä, vaan vain niistä verbeistä, joita aineistoon on osunut. Määrälliset tulokset voivat siis olla korkeintaan suuntaa antavia.

Käsitellessämme aineistossa esiintyviä ilmiöitä laadullisesti olemme ottaneet mukaan liitteenä 2 olevan listan ulkopuolelta myös joitakin sellaisia verbejä, jotka valottavat *U*-verbeihin liittyviä kiinnostavia oppimisen ilmiöitä. Monet näistä kertovat luovasta sanojen johtamisesta (esim. **murtoudui pro murtui*). Akateemisten kirjoittajien LAS2-aineistossa esiintyy myös uudismuodosteita (esim. *orientoidaan*), kun kirjoittajat pyrkivät mahdollisimman tieteelliseen ja kompleksiseen kielenkäyttöön.

6. *U*-verbien esiintyminen ja käytön tarkkuus

Tarkastelemme tässä luvussa edellä esiteltyjen verbien määriä eri taitotasoilla ja kolmessa eri korpuksessa. Esittelemme myös, minkälaisia virheitä näiden verbien käytössä on eri tasoilla. Seuraavissa taulukoissa (4–7) on etsittyjen verbien yhteismäärät taitotasoittain ja aineistoittan. Taulukoissa ei ole eroteltu *A*- ja *U*-verbejä, koska tämän artikkelin teema on juuri näiden kahden ryhmän mahdollinen sekaantuminen oppijankielessä. Jäljempänä on esimerkkejä näiden kahden ryhmän määrällisistä ja laadullisista suhteista. Joissakin verbipareissa *A*-verbi on yleinen, joissakin *U*-verbi. Verbikohtaisia lukuja ja verbien käyttöä on esitelty kunkin taulukon yhteydessä.

TAULUKKO 4. *Aineistojen A- ja U-verbien esiintymät tuhatta sanetta kohti sekä eri verbien määrät*

Esiintymät	Taitotaso			
	Kouluaineisto	YKI-aineisto	LAS2-aineisto	Yhteensä
Esiintymiä	240	568	1464	2272
/1000 sanaa	9,2	10,7	14,2	12,5
Eri verbejä	43	59	58	78

Taulukkoon 4 on koottu kaikkien aineistojen kokonaiskuva. Luvut eivät tietenkään ole vertailukelpoisia, sillä vain YKI-aineisto edustaa kaikkia taitotasoja. Kouluaineisto edustaa vain taitotasoja A1–B2, ja LAS2-aineiston kaikki tekstit ovat vähintään tasoa B2. Taulukoissa 5–7 on siksi esitelty kunkin aineiston määrälliset tunnusluvut erikseen taitotasoittain (koulu- ja YKI-aineisto) tai tekstilajeittain (LAS2-aineisto, ks. luku 5).

TAULUKKO 5. *Kouluaineiston A- ja U-verbien esiintymät tuhatta sanetta kohti sekä eri verbien määrät taitotasoittain*

Esiintymät	Taitotaso						
	A1	A2	A yht.	B1	B2	B yht.	Kaikki
Esiintymiä	22	74	96	117	27	144	240
/1000 sanaa	6,3	8,2	7,7	10,3	12,1	10,6	9,2
Eri verbejä	11	21	23	30	15	33	43

Kouluaineistossa verbien esiintymien yhteismäärä tuhatta sanetta kohti kasvaa tasaisesti tasolta A1 tasolle B2. Tason B2 tekstien määrä kouluaineistossa on kuitenkin niin pieni, ettei siitä voi juuri tehdä päätelmiä. Aineiston pienuus näkyy varsinkin eri verbien määrässä. Tässä tutkittujen verbien käyttö kuitenkin lisääntyy tasolta toiselle, eli ne saavat suuremman osuuden koko sanamäärästä. Tämä voi kertoa yhtä hyvin juuri näiden verbien käytön lisääntymisestä kuin siitäkin, että alimmilla tasoilla käytetään yleensäkin hyvin pientä valikoimaa sanoja ja etenkin

verbejä (ks. esim. Puro 2002, jossa todettiin yhden alkeiskurssin osallistujien käyttävän aktiivisesti ensimmäisen lukukauden lopussa vain 9:ää eri verbiä). Sama verbien käytön yleinen lisääntyminen näkyy siinä, että tässä käsiteltävien eri verbien määrä kasvaa tasolta A1 tasolle B1 11:stä 30:een.

Liitteen 2 verbeistä yleisimmät kouluaineistossa ovat *nähdä* (49 esiintymää), *vaihtaa* (38), *kuulla* (22), *muistaa* (18), *unohtaa* (16), *näyttää* (13), *kaatua* (10) ja *näkyä* (7). Näistä *vaihtaa* liittyy selkeästi yhteen tehtävään, verkko-ostoksesta valittamiseen. Tehtävänannossa pyydettiin kertomaan, miten toivoisi asian hoidettavaksi, ja moni ehdotti tietokonepelin vaihtamista. Muut verbit esiintyvät eri tesktilajeissa tasaisemmin. Alla on esimerkkejä (1–8) niiden tyypillisistä käytöistä. Esimerkkien kirjoitusasu välimerkkeineen on alkuperäinen.

- (1) mutta pahin on viimeks yöllä **nähtiin** joku eläin kulji metsässä, ja sitten juoksin pois sieltä Eläi **en ole nähnyt** on kyy (A1-F2017T5)
- (2) Haluaisin että siihen **vaihdetaan** uusi kalvo koska se on naarmuuntunut paljon. (A2-F3007T3)
- (3) **Kuulin**, että meillä on joku koe ensi viikolla. (B1-F7608T2)
- (4) hän opetti sen niin hyvin minulle että minä **muistan** sen vieläkin (A2-F0557T5)
- (5) Tänään kun minä olin menossa bussi pysäkillä huomasin että **olen unohdanut** minun loppakoni kotiin. (A1-F7448T2)
- (6) Se peli **näyttää** aika kivalta mutta se ei toimii. (B1-F7308T3)
- (7) Tapahtuma oli hauska koska kun se **kaatui** se oli naurettava... (A2-F7028T5)
- (8) silloin kun saan pisteitä, ruudussa **ei näy**, että sain pisteen (B1-F7889T3)

YKI-aineistossa tutkittujen verbien käyttö lisääntyy vielä selvemmin taitotasolta toiselle, kuten havaitaan taulukosta 6.

TAULUKKO 6. YKI-aineiston A- ja U-verbien esiintymät tuhatta sanetta kohti sekä eri verbien määrät taitotasoittain

Esiintymät	Taitotaso									
	A1	A2	A yht.	B1	B2	B yht.	C1	C2	C yht.	Kaikki
Esiintymiä	28	33	61	84	92	176	149	182	331	568
/1000 sanaa	5,6	5,8	5,7	7,7	10,1	8,8	12,9	16,8	14,8	10,7
Eri verbejä	10	16	17	26	31	40	35	39	49	59

Kiinnostavaa on, että B-tasolla käytön yhteismäärä on noin kolmanneksen suurempi kuin A-tasolla, mutta C-tasolla taas liki kaksinkertainen verrattuna B-tasoon. Tutkitut verbit ovat siis ominaisempia ylimpien tasojen kirjoitelmille. Eri verbien määrä ei lisääny aivan samaan tapaan, vaikka niitäkin on eniten C-tasolla, vaan valikoima lisääntyy tasaisemmin A1-tason 10:stä C2-tason 39:ään. Eri verbien määrää rajaavat tietenkin tehtävänannot, joita YKI-kokeissa on käytetty. Ne ovat sisällöltään vaihtelevampia kuin kouluaineiston tehtävänannot, jotka ovat kaikille samat (ks. luku 5).

Tavallisimmat verbit YKI-aineistossa ovat *järjestää* (53 esiintymää), *vaihtaa* (43), *muistaa* (39), *nähdä* (37), *löytää* (25), *muuttaa* (25), *unohtaa* (21), *liittyä* (20), *näyttää* (19), *tuntua* (19), *kuulla* (18) ja *kuulua* (17). Yhtenä tehtävänä oli talkookutsun kirjoittaminen, ja *järjestää*-verbejä esiintyy erityisesti näissä teksteissä. *Vaihtaa* on myös aikuisilla tavallinen reklamaatioissa ja *liittyä* taas tehtävässä, jossa piti kysellä jonkin järjestyksen toiminnasta.

Esimerkkejä (9–19) eri tasojen aikuisten YKI-teksteistä:

- (9) Taloyhtiömme **järjestää** perinteiset kevättalkoot sunnuntaina 28.4. (C1-F283CT1)
- (10) mutta haluisimme **vaihtaa** opetusaikaa (A2-F609AT3)
- (11) minä ihan **unoidin** että minä velkaan sinulle rahaa, ja **muistin** vaan tanaan kun äiti antoi minulle rahaa. (A1-F353AT2)

- (12) Ystävä, joka **en nähnyt** pari vuotta tulee (A2-F362AT2)
 (13) Kaupungissa aina **löydät** jotakin tekemistä (B1-F656BT3)
 (14) Sääli, että sinä **muuttanut** ulkomaille. (B1-F343BT2)
 (15) Haluan selvittää, millä ehdolla henkilö pystyy **liittymään** yhdistykseen. (C2-F1007CT1)
 (16) Miltä esim ke iltapäivä **näyttää?** (B2-F703BT1)
 (17) Välillä **tuntuu**, että on hyvin vaikeaa saada aikaa lääkärille vaikka tilanne olisi vakavaa. (C1-F236CT3)
 (18) Olen **kuullut** että Henkilökuntayhdistykseen liittyminen on nyt maksuton. (C2-F1009CT1)
 (19) Lapsen menestyminen aikuisena ja hyvä kasvatusta **kuuluvat** yhteen. (B2-F680BT3)

TAULUKKO 7. LAS2-aineiston A- ja U-verbien esiintymät tuhatta sanetta kohti sekä eri verbien määrät tekstilajeittain

Esiintymät	Taitotaso			
	Tenttivastaukset	Esseet	Tutkielmaluonnokset	Kaikki
Esiintymiä	572	312	580	1464
/1000 sanaa	17,4	16,0	11,4	14,2
Eri verbejä	51	47	44	58

Edistyneiden oppijoiden akateemisten tekstien aineistossa (taulukko 7) ei ole kyse taitotasojen vaan tekstilajien eroista. Tenttivastauksissa ja esseissä on eniten tässä haettuja verbejä, tutkielmaluonnoksissa hieman vähemmän. Kaiken kaikkiaan näiden verbien esiintymiä on enemmän (14,2/1000) kuin kouluaineistossa (9,2) ja YKI-aineistossa (10,7). Eri verbien määrä (58) on tässä akateemisessa aineistossa jokuinkin sama kuin aikuisten yleiskielitaitoa edustavassa YKI-aineistossa kaiken kaikkiaan (59), mutta YKI-aineistossa kokonaismäärä koostuu eri tasoilla esiintyvistä erilaisista verbeistä siten, että A1-tasolla eri verbejä on 10, C2-tasolla 39. Tässä haettujen verbien variaatio on siis selvästi suurempi akateemisessa aineistossa, vaikka siinäkin on mukana hyvin tavallisia ja konkreettisia verbejä.

Tenttiaineistossa on eniten eri verbejä, ja näitä verbejä on myös käytetty eniten tuhatta sanetta kohti. Tämä selittyy sillä, että tentti-vastausten aihepiirit ovat hyvin moninaiset ja vastauksia on useammilta informanteilta kuin esseitä ja tutkielmaluonnoksia, jotka sitä paitsi keskittyvät kukin yhteen aihepiiriin.

Edistyneiden oppijoiden aineiston tavallisimmat verbit ovat *kuulua* (160 esiintymää), *liittyä* (134), *muodostaa* (94), *kehittyä* (91), *muuttua* (89), *nähdä* (77), *näkyä* (63), *aiheuttaa* (60), *löytyä* (52) ja *johtua* (44). Alla on esimerkkejä (20–29) näiden verbien käytöstä LAS2-aineistossa.

- (20) Kantakielen tempussysteemiin **kuuluivat** preesens ja yksi mennyt aika. (las2-6tt03te01)
- (21) Kansallispuvun syrjäytyminen **liittyi** talouselämän muutoksiin. (las2-7tt03te01loxx)
- (22) Tämän hypoteesin mukaan olisi joka tapauksessa oletettavaa, että niin suomen kielen puhujille tärkeä ja vanha perinne kuin sauna **muodotaisi** olennaisen osan sanastoa. (las2-3tt03te02lo04)
- (23) Suomen demonstratiivipronomini *se* on **kehittymässä** artikkeliksi. (las2-1tt01te07lo13)
- (24) Kun vieraat rikkovat tavallisen kysymys–vastaus-kuvion, keskustelun kulku **muuttuu** arvaamattomaksi. (las2-5tt02te01lo19)
- (25) Nuorien lainojen tapauksissa **on nähtävissä**, (las2-4tt01te05lo03)
- (26) Kieliopillistuminen **näkyä** eniten puhekielessä (las2-1tt01te07lo13)
- (27) Virheet voivat olla --- itse opetusprosesstin **aiheuttamia** virheitä. (las2-9tt01te01lo00)
- (28) Verbistä *sivellä* muodostettiin 13 kertaa *sivelen* ja 7 kertaa *sipelen*, johon selitys voi **löytyä** jälleen sanan tuntemattomuudesta (las2-2tt02te05lo20)
- (29) Vieraantuneisuus **johtune**e enemmän taloudellisista ongelmista ja alemmuudentunteista kuin ympäristönmuutoksesta. (las2-3tt03te03lo20)

Edeltä käy ilmi, että eri aineistojen tavallisimmat verbit ovat osin samoja, osin aivan erilaisia. Kaikissa yleinen on *nähdä*-verbi. Sen merkitys koululaisilla ja YKI-aikuisilla on useimmiten konkreettinen ja viittaa joko visuaaliseen havaintoon tai tapaamiseen. Toista aistivahaintoverbiä

kuulla on LAS2-aineistossa melko vähän (17 esiintymää), kun taas sen pari *kuulua* on kyseisessä aineistossa eniten esiintyvä tässä puheena oleva verbi. Kiinnostavin ero aineistojen välillä on se, että kouluaineistossa tavallisimpien verbien joukossa on *U*-verbejä vain kaksi (*kaatua* ja *näkyä*), YKI-aineistossa kolme (*liittyä*, *tuntua* ja *kuulua*), kun taas LAS2-aineiston kymmenen yleisimmän verbin joukossa on peräti seitsemän *U*-verbiä.

Aineistosta haettiin yhteensä 84:ää eri verbiä. Näistä seuraaville kuudelle ei löytynyt yhtään esiintymää: *myöntyä*, *naarmuttaa*, *paistua*, *peittyä*, *pelastua* ja *peruuttaa*. Tämä johtuu siitä, että kun aineistosta tätä tutkimusta varten haettiin *A*- ja *U*-verbejä, lisättiin listaan kunkin löydetyn verbin pari. Kaikkien parien molempia jäseniä ei kuitenkaan todellisuudessa aineistossa siis ollut.

Virheiden määrää ei tässä ole taulukoitu, sillä virheitä *A*- ja *U*-verbin valinnassa tai käytössä on kaiken kaikkiaan hyvin vähän, vaikka alimmilla taitotasoilla on toki monenlaisia muita virheitä niin oikeinkirjoituksessa, taivutuksessa kuin lauseenrakentamisessakin. *U*-verbejä käytetään kuitenkin alimmilla taitotasoilla harvoin ja silloinkin usein melko vakiintuneissa konteksteissa, joissa oppija ei päädy pohtimaan valintaa näiden kahden verbiryhmän välillä. Joskus (30–32) kuitenkin tarvitaan verbiä, jonka käyttöä ei hallita:

- (30) Salmiakki – se **maistaa** pahalta (A1-YKI-F268AT3)
- (31) en voi tulla tunnille koska olen sairas, voiko **siirtyä** se koe huomiseen (A2-koulu-F7598T2)
- (32) Tallä hetkellä vaimoni ja lapseni ovat sairas, tavulinen flunsa. Toivon **parantavat** piian. (B1-YKI-F326BT2)

YKI-aineiston ylimmilläkin tasoilla *A*- ja *U*-verbin valintavirheet (33–34) ovat varsin harvinaisia, mutta niitä on etenkin melko mutkikkaissa rakenteissa:

- (33) Meillä on pitkään ollut käsitys, että suomen terveydenhuolto on maailman parhaita, mutta käsitykseni **muutti** täysin, kun luin OECD:n selvityksen terveydenhuollosta eri maissa. (C1-YKI-F247CT3)

- (34) En voi luottaa näin välinpitämättömään lääkäriin, joten pyydän ystävällisesti **vaihtua** toiselle omalääkärille (C2-F1144CT2)

Esimerkissä (34) on myös mahdollista, että kyseessä ei ole verbivalinta vaan sanahahmojen sekaantuminen tai murteellisuus, sillä joillakin alueille *vaihto*-sanana partitiivi voisi olla *vaihtua*.

Ilmeisesti kuitenkin tekstin yleinen kompleksisuuden kasvu lisää virheellisiä *A*- ja *U*-verbivalintoja, sillä eniten on ongelmia LAS2-aineiston akateemisissa teksteissä. Alla on muutamia esimerkkejä (35–38) näistä:

- (35) että vanha* e on **kehittänyt** i: ksi sanan toisessa tavussa. (las2-8tt01te04lo04)
(36) kantasana voi **muuttaa** tai kadota (las2-7tt01te02lo01)
(37) Lisäksi Inkerinmaalle ja Käkisalmen läänille alkoi **muuttua** Ruotsin valtakunnan asukkaita, siis suomalaisia, (las2-5tt01te04lo05)
(38) Toisin sanoen vainajan sukulaiset **valmistuivat** tavarat niin, (las2-8tt01te09lo18)

Nämä esimerkit ovat tenttivastauksista, joissa tarve osoittaa sekä opittavan sisällön tuntemusta että suomen kielen hallintaa kilpailee aikapaineen kanssa. Erilaiset sanahahmo-ongelmat ovat tavallisia myös LAS2-aineistossa. Johtimen *U* sekoittuu helposti taivutusainesten *U*:n kanssa, jolloin syntyy sellaisia muotoja kuin **kehytetty* (pro *kehitetty/kehittynyt*) tai **jatkonut* (pro *jatkunut*).

Seuraavassa luvussa tarkastelemme erityisesti *U*-verbien ja niiden *A*:llisten parien käytön ilmiöitä ja kehitystä eri taitotasoilla. Luvussa 8 palaamme pohtimaan tulosten mahdollisia selityksiä ja niiden merkitystä taitotason indikaattoreina.

7. Kompleksisuuden kehitys

Kompleksisuuden käsite toisen kielen oppimisen alalla on pian jo parin vuosikymmenen ajan ollut jatkuvan kiinnostuksen kohteena. Monet pitävät sitä puhtaasti laskennallisena suureena ja käyttävät sen mittarina esimerkiksi ilmausten morfeemimäärää (MLU = *Mean Length of Utterance*) tai alisteisten lauseiden määrää päälauseetta kohti. Monia muita-

kin mittareita on kokeiltu (ks. esim. Wolfe-Quintero ym. 1998). Kaikki mittarit ovat kuitenkin tavalla tai toisella ongelmallisia. Kirjoitetussaakaan oppija-aineistossa ei ole aina mahdollista jakaa tekstiä lauseiksi, puheesta puhumattakaan. Eräässä kokeessa joukko suomen kielen opiskelijoita pantiin jakamaan oppijoiden tekstejä lauseiksi. Näin löydettyjen lauseiden määrä samassa testissä vaihteli jopa 35–48. (Martin tulossa). Alisteisuuskaan ei ole suomen kielen osalta välttämättä hyvä kompleksisuuden mittari, sillä vaikka alisteisuus on varmasti kognitiivisesti rinnastamista kompleksisempaa ja voi siksi toimia ensikielen kompleksisuuden mittarina, kaikki tämän aineiston kirjoittajat osaavat alistuksen ensikielessään. Suomen alisteiset sivulauseet eivät ole erityisen hankalia, sillä relatiivilauseita lukuun ottamatta niihin ei liity päälauseista poikkeavia ongelmia. Siksi tässä kompleksisuutta ei pidetä määrällisenä vaan laadullisena käsitteenä. Tämän lähestymistavan puolesta on argumentoinut laajemmin esimerkiksi Reiman (2011).

Kompleksisuuden laadullinen määrittely sisältää distribuution käsitteen (ks. myös Martin ym. 2010). Distribuutiolla tarkoitetaan tässä niitä konteksteja, joissa jokin sana tai rakenne – tässä siis *U*-verbin sisältävä konstruktio – esiintyy. Tarkastelemme alla muutamia esimerkkitapauksia edellisen luvun määrällisen analyysin pohjalta. Liitteen 2 verbeistä kaikissa aineistoissa yleisiä ovat esimerkiksi *kuulla/kuulua*, *nähdä/näkyä*, *löytää/löytyä*, *muuttaa/muuttua* ja *liittää/liittyä*.

7.1. *kuulla/kuulua*

- (39) **Olen kuullu** et suomen kielen koe on tulossa pian (A2-koulu-F0427T2)
- (40) ei edes pikku ääntä **kuuluu**, eika, pienet valot palaa (B1-YKI-F664BT2)
- (41) Perustetiin kouluja tai mahdollisuuksia liiviläisille lapsille **kuulla** koulussa myös liiviä (las2-2tt01te03lo04)
- (42) “korvalla” ilmaisee paikkaa, läheisyyttä; ehkä sillä on myös muut konotaatiot niin lähellä, että voi **kuulla**, mitä toinen puhuu, että isä ja lauseen kertoja olivat paikalla yhdessä, ei yksin (las2-2tt03te06loxx)

- (43) Mutta kieleen **kuuluu** muutakin kuin ymmärrettävyys. (las2-2tt03te10lo21)
- (44) Samanaikaisia Muualta[sic] Euroopasta tulevista uhkakuvista huolimatta tämä toimi hyvin, ehkä sen takia, että saunaan on **kuullut** tiettyjä kirjoittamattomia sääntöjä, jotka pitävät saunan hiljaisena ja rauhallisena paikkana. (las2-3tt03te02lo04)
- (45) Voi aloittaa siitä, että ensin selittää, että tietynlaiset sanat vaativat niiden kanssa **kuulevissa** sanoissa tietynlaisia päätteitä, esim. puhun suome-a, kiitos kahvi-sta jne. (las2-5tt01te02lo02)

Alemmilla taitotasoilla *kuulla* esiintyy muutaman kerran konkreettisesti aistihavainnon merkityksessä, liki kaikki *kuulua*-verbin esimerkit taas ovat fraasin *mitä kuuluu* -osia, joita ei tässä ole laskettu mukaan. LAS2-aineistossa esimerkit sisältävät pääasiassa johonkin ilmiöön tai kieliryhmään kuulumisen. Esimerkeissä (44) ja (45) on johonkin kuulumista kuvaavissa ilmauksissa kuitenkin käytetty *kuulla*-verbin muotoja. Partisiippi *kuullut* saattaa oppijalle hahmottua helpommin verbin *kuuluu* kuin *kuulee* muodoksi. Hämätä voi myös se, että johonkin kuulumisella ei ole semanttista transitiivista paria.

7.2. *nähdä/näkyä*

- (46) Eläi **en ole nähnyt** on kyy (A1-koulu-F2017T5)
- (47) Yksi ongelma on se että **ei näy** mitään muuta kanavaa paitsi ykkönen ja kakkonen. (B2-YKI-F698BT2)
- (48) Vainajia **nähdään** [pidetään] myös neuvonantajina ja moraalin valvojina. (las2-4tt01te09lo21)
- (49) Täysin kieliopillistuneessa postpositiossa sanan alkuperää **ei** enää **näy**. (las2-3tt01te04lo16)

Myös *nähdä* ja *näkyä* ovat alemmilla tasoilla konkreettisia, kun taas akateemisessa tekstissä ne muuttuvat abstrakteiksi. Samalla myös lauseyhteydet ovat kieliopillisesti mutkikkaampia. Verbien *nähdä* ja *näkyä* keskinäinen suhde on hyvin samanlainen kuin seuraavassa olevien verbien *löytää* ja *löytyä*. Niissä molemmissa ilmiötä tarkastellaan kokevan agentin suunnasta ja eron oivaltaminen saattaa tuottaa vaikeuksia.

7.3. löytää/löytyä

- (50) sitten mun ystäväni **löysi** [puhelimien] ja se antoi mulle sitte mulla tuli ilo. (A1-koulu-F7758T3)
- (51) Paketistä **löytyivät** teidän yhteystiedot, palautankö samaan osoiteen (B2-YKI-F678BT2)
- (52) Tutkimuksessa pyrin **löytämään** ja analysoimaan niitä strategioita, joita saman etnisen ryhmän suomen oppijat valitsevat sanavalinnassa. (las2-1tt02te01lo18)
- (53) Sanalle **löytyy** vastineita indoeurooppalaisista kielistä. (las2-4tt01te05lo03)

7.4. muuttaa/muuttua

- (54) Minä haluan. **muuta** uutta peli. Jos voi **muuta** soita minulle. (A1-koulu-7819T3)
- (55) Vasta 3-kuukautta sitten sinä **muuttaisit** Ruotsiin, ja nyt sinulla on häätilaisuus. (B1-YKI-F332BT2)
- (56) että sanajärjestys lauseessa voi **muuttaa** merkitystä (las2-4tt01te06lo02)
- (57) ja monen asenne **on muuttunut** myönteisemmäksi. (las2-3tt01te07lo1)

Esimerkissä (54) *muuttaa* on merkityksessä *vaihtaa*. Ainakin balttilaisissa ja useissa slaavilaisissa kielissä sekä myös englannissa käytetään samaa verbiä ilmaisemaan näitä kahta ilmiötä (Siitonen 1999: 386–387). Esimerkissä (55) taas on kyse konkreettisesta asunnosta toiseen muuttamisen merkityksestä. Esimerkeissä (56) ja (57) verbin merkitys on abstrakti.

7.5. liittää/liittyä

- (58) Haluaisin **liittyä** “Helsingin matkaoppaat ry”:n jäseneksi. (C2-YKI-F1001CT1)
- (59) 1900-luvulla persoona- ja demonstratiivipronominien lisäksi deiksikseen **liitettiin** myös aikaa ilmaisevia adverbeja ja tempukset. (las2-6tt02te02lo29)

- (60) Ensikielen omaksuminen **liittyy** kielenoppijan yleiseen kehitykseen. (las2-1tt02te01lo18)

Verbiä *liittää* tai *liittyä* ei esiinny lainkaan alemmilla kielitaidon tasoilla. Vasta C-tasolla se ilmaantuu kirjoitelmiin muun muassa merkityksessä *liittyä yhdistykseen*. Akateemisissa teksteissä sen sijaan *liittää* on melko tavallinen (14 esiintymää) ja *liittyä* kaikkein yleisimpien verbien joukossa (59 esiintymää).

7.6. *unohtaa/unohtua*

- (61) Kirjoitat sinulle, koska sinä **olet unohtanut** kellon luokseni. (A1-YKI-F1118AT2)
- (62) Kunhan [poliitikko] vaan pääsee parempaan paikkaan ja lupaukset saa **unohtaa**. (B1-YKI-F416BT3)
- (63) Kansa **unohtuu** ja kaikki luvattut asiat jäävät tekemättä. (B1-YKI-F403BT3)
- (64) näin ollen puvun henkilökohtainen viesti **unohtuu**. (las2-5tt01te07lo18)

Alimmilla tasoilla unohdetaan esineitä, myöhemmin lupauksia tms. Ylemmillä tasoilla käytetään verbiä *unohtua* osuvasti häivyttämään unohtaja, jota ei välttämättä pystyttäisikään yksilöimään.

7.7. Kielitaito kehittyy – virheet lisääntyvät

Mitä korkeammalla tasolla oppijoiden kielitaito on, sitä monimutkaisemmiksi *U*-verbi-ongelmat muuntuvat. LAS2-aineiston informantit ovat lukeneet runsaasti käsitteellistä ja nimenomaan kielitieteellistä kirjallisuutta. Tietynlaisen tekstin lukeminen ja erityisesti tutkielman lähtenä käyttäminen väistämättä vaikuttavat omaan kirjoittamiseen. Tenttivastauksissaan, esseissään ja tutkielmissaan informantit pyrkivätkin kirjoittamaan kyseiseen tekstilajiin kuuluvaa kieltä. Kirjoittajat käyttävät paljon tieteellisissä teksteissä suosittuja infiniittisiä rakenteita. Koska kirjoittajalla ei ole kyseessä kielitesti vaan itse tutkitun asian esille

saattaminen, hän ei ensisijaisesti kiinnitä niin runsaasti huomiota korrektiuteen kuin asiasisältöön.

- (65) Kirjoituksesta **johtavaa** erehtymistä ei ole ollut myös sanan öhtu tai vutuksessa, öhtu on astevaihteluton, yksi vastaus oli öhu, jossa voi nähdä oikean kirjoitusasun tulkinnan. (las2-2tt02te05lo20)
- (66) Normaali (suora) astevaihtelu tapahtuu yleensä suffiksien **liittyessä** (esim. takki : taki-ssa), käänteinen astevaihtelu voi tulla kyseeseen vokaalivartalon muodostaessa (esim. kirkas : kirkkaa-). (las2-2tt02te04lo19)
- (67) Verbien taivutusmuodoilla on (yksikön imperatiivia ja preesensin kieltoa lukuun ottamatta) aina pääte, taivutus tapahtuu sekä päätteen **liittymällä** että vartalon asteen vaihdolla: (las2-2tt02te01lo18)

Esimerkissä (65) on tyypillinen *A-U*-ongelma, ja lisäksi virkkeen pituus ja *myös*-sanon käyttäminen *myöskään*-sanon sijasta tekee virkkeestä hyvin vaikeasti ymmärrettävän. Esimerkeissä (66) ja (67) ei ole aivan selvää *A-U*-ongelmaa: suffiksit ja päätteet *liittyvät*, vaikkakin yleensä puhutaan niiden *liittämisestä* tai ainakin mainitaan, mihin ne *liittyvät*. *U*-verbin esiintyminen lausekokonaisuudessa on lisäksi epäsymmetristä, koska *U*-verbeihin rinnastetaan *A*:lliset ilmaukset *muodostaessa*, jonka sijaan olisi vielä parempi käyttää passiivista *muodostettaessa*, ja *vaihto*, joka on johdettu *vaihtaa*-verbistä.

Agenttiuden ilmaiseminen idiomaattisella tavalla ei ole helppoa: oppijan oman ensikielen vaikutus, sarjojen aukkoisuus ja intentionaalisuuden moninainen ilmaiseminen suomen kielessä sekä ilmisubjektittomien ilmausten – niin passiivin kuin nollapersoonankin – viittausten ja sävyjen hallinta tuottaa ongelmia (vrt. edellä luku 4). LAS2-aineiston kirjoittajat hallitsevat periaatteessa kaikki eri keinot kieliopillisesti. Kyse on kuitenkin siitä, mikä ilmaus kussakin tilanne- ja tekstikontekstissa olisi viisainta valita.

Esimerkeissä (68) ja (69) on kyllä *A-U*-suhde kohdallaan, mutta taivutuksen monimutkaisuuden ilmeneminen (68) ilmaistuna passiivilla, jonka taustalla olisi ihmispersoonaa, on outoa, samoin kieltenväliset typologiset kysymykset (69) näin ilmaistuina.

- (68) Testissä ei ollut sanoja, joiden taivutuksen monimutkaisuus on **aiheutettu** paitsi astevaihtelulla muilla vartalon vaihtelulla, esim. verbitaivutus tyyppiä kahelta : kahtlen tai vielä monimutkaisemmat verbitaivutukset kuten andma : anda : annan : antakse, petma : petta : petan : petetakse. (las2-2tt02te05lo20)
- (69) Tämä on **heijastettu** myös siinä, että suomea ja viroa pidetään typologisesti erilaisina (esim. vain virossa voi astevaihtelun allaistenallomorfiien vaihtelu yksin kantaa morfologisten kategorioiden opposition). (las2-2tt02te06lo20)

Aivan päinvastaisesta tapauksesta on kyse esimerkissä (70), jossa on luontevasti käytetty *U*-verbiä mutta johon idiomaattisempi olisikin passiivimuoto.

- (70) Perfekti **muodostuu** II partisiipilla ja olla-verbin avulla. (las2-7tt01te04lo03)

Samoin esimerkissä (71) passiivi olisi tyyppillinen ratkaisu. Ilmaukselle *perustaa seura* ei ole vastaavaa *U*-muodostetta pariksi; kyseessä on aukko sarjassa. Ilmaus *seura perustui* ei kuitenkaan ole kovin kaukana ilmauksista *asia hoitui* ja *halli rakentui*.

- (71) 1830-luvulla **perustui** Lauantaiseura. (las2-8tt01te06lo04)

Nollapersoonan käyttö vaikuttaa ensi näkemältä varsin korrektilta esimerkissä (72). Inessiivimuoto ei kuitenkaan ole aivan luonteva; elatiivi sopisi nollapersoonaan paremmin. Todennäköinen ensikielisen suomenpuhujan ratkaisu olisi tässä *ei näy*.

- (72) Näiden sanojen kohdalla, joissa kirjoituksessa asteen eroa **ei näe**, ilmoittavat luvut, kuinka paljon informanttia piti sanaa astevaihtelun alaisena ja kuinka paljon astevaihtelun ulkopuolella. (las2-2tt02te01lo18)

Nollapersoonaa edellyttää taustalle jotakuta elliptiseksi tutkijaksi, huomaajaksi tai katsojaksi. Esimerkissä (73) kuitenkin näyttää olevan kyse kielen ilmiöstä, joka näkyy tai ei näy ja jonka yhteydessä katsojaa ei korosteta.

- (73) Sen kongruenssipiirre **voi nähdä** esim. omistusrakenteessa. (las2-8tt01te08lo14)

Esimerkissä (73) on myös kyse nollapersoonalauseesta, jossa on käytetty inessiivimuotoa. Lisäksi objekti on nominatiivissa. Tämä voidaan tulkita kahdella tavalla: kirjoittaja on pyrkinyt kirjoittamaan nollapersoonalauseen tai sitten kyseessä onkin A-U-erehdys (*nähdä pro näkyä*). Kirjoittajan tarkoitusperiä ei voi virkkeen perusteella saada selville. Vastaavanlaista, mutta päinvastaista tapausta on Anna Denisyuk (2010: 80) käsitellyt pro gradu -tutkielmassaan. Denisyuk kommentoi sanavalintaa esimerkissä --- *siis historiaalis-vertailevan kielitieteen avulla voi näkyä suomalais-ugrialaisten kielten sukulaisuuden seuraavasti*: “Jos sana *sukulaisuus* olisi nominatiivissa, sanavalinta *näkyä* olisi oikea. Informantti kuitenkin valitsi objektin sijan, mikä voi tarkoittaa, että tavoiteltu sana oli *nähdä*.”

Näyttää siltä, että akateemisella tasolla nimenomaan agenttisuhteet tuottavat pääasialliset ongelmat. Ongelmalliset agenttisuhteet eivät rajoitu pelkästään U-verbeihin. Transitiivisuus–intransitiivisuus-sekaantumia, joissa ei tekijää ehkä ole ollenkaan, esiintyy myös muilla verbeillä. Esimerkissä (74) sekä pari *jättää–jäädä* että pari *vahvistaa–vahveta* ovat aiheuttaneet kirjoittajalle valintaongelmia.

- (74) Vahveneeko vai heikkeneeko aste voi olla yksi tekijä, joka vaikuttaa tuloksiin, vaihtelua voi ehkä helpommin **jättää vahvenematta**, tämä todistaisi myös verbin süüdata : süütan (3:19) taivutus, jossa 13 kertaa on ollut muodostettu 1. persoona süüdan. (las2-2tt02te05lo20)

Agentin ilmaisu, agentin intentio, agentin poissaolo sekä lopulta kirjoittajan intentio kutoutuvat yhteen erityisesti kompleksisissa infiniittisissä rakenteissa. Yksinkertaisen ilmauksen tuottaminen onnistuu, mutta ilmeisesti useiden asioiden kasaantuessa yhteen muisti kuormittuu liikaa ja ilmauksen rakentamisesta ei enää selviydytä.

8. Pohdintaa

Tämän artikkelin ja sen taustalla olevan tutkimushankkeen keskeinen ajatus on etsiä lingvistisiä ilmiöitä, joiden esiintyvyys, hallinta tai käyttötapa eroavat eri taitotasolla, kun tarkastellaan suurta joukkoa tiettyä funktionaalista taitotasoa edustavia tekstejä. Tässä tarkastelun kohteeksi on valittu *A*- ja *U*-verbiparit, joista 42:ta paria on haettu yhteensä noin 180 000 sanan aineistosta. Korpuslingvistiseksi aineistoksi se on tietenkin pieni, mutta suomen kielen oppijoiden kirjoitelma-aineistoja, joiden tehtävänannot ja taitotasot ovat tiedossa, ei ole olemassa vielä kovin paljon. Aineiston pienuudesta huolimatta tulokset ovat kiinnostavia. Sekä kirjoittajien taitotaso että tekstilaji vaikuttavat tutkittujen verbien esiintymistiheyteen ja käytön laatuun. Lisäksi yksittäisten verbien taajuuteen vaikuttaa kirjoitelman aihepiiri: *naarmuuntua* olisi tuskin valikoitunut tähän tutkimukseen ilman tehtävänantoa, jossa piti valittaa nettikauppaan tietokonepelistä, *muuttua* taas on sidoksissa kielihistoriallisiin aihepiireihin.

Tehtävänannon, aihepiirin ja kirjoittajan yksilöllisten ominaisuuksien lisäksi myös lähdekielellä on vaikutusta kirjoittajan ratkaisuihin. Erityisesti näin näyttää olevan silloin, kun suomen kielen opinnot on aloitettu Suomen ulkopuolella, kuten LAS2-aineiston maisteriopiskelijoista suurin osa on tehnyt. Suomea on opiskeltu ensikieleltään homogeenisessa ryhmässä vuosikausia, ja usein opetuskielenäkin toimivan ensikielen vaikutuksella on ollut tilaisuus vakiintua. Vertailu Oulun ICLFI-korpukseen saattaisi tuoda kiintoisia tuloksia.

Tarkasteltujen verbien käyttö lisääntyy kohti ylempiä taitotasoa ja vaativampia tekstilajeja. Koululaisten teksteissä niiden käyttö lisääntyy *A*- ja *B*-tason välillä 7,7:stä 10,6:een tuhatta sanetta kohti. Aikuisten yleisluontoisissa teksteissä vastaavat luvut ovat 5,7 tasolla *A*, 8,8 tasolla *B* ja 14,8 tasolla *C*. Akateemisissa tekstilajeissa, jotka edustavat enimmäkseen tasoa *C*, määrä on 14,2. Ylimpien tasojen tekstit eivät siis suuresti eroa toisistaan määrällisesti, mutta verbivalinnoissa on eroa. Akateemisissa teksteissä *U*-verbejä on selvästi enemmän, kun taas muissa

teksteissä on enemmän *A*-verbejä. *U*-verbien taajuus siis on ylimpien taitotasojen indikaattori.

Tutkittavan ilmiön hallinta eli virheiden määrä voi esiintymistajuuden ohella vaihdella taitotasolta toiselle. Useimmiten ajatellaan, että hallinta paranee siirryttäessä kohti ylempiä taitotasoja. Näin onkin, kun tarkastellaan suurta joukkoa ilmiöitä, ja saattaa olla jopa niin, että ilmiön esiintymistajuuden ja hallinnan välillä vallitsee melko säännöllinen suhde. Useimpien ilmiöiden taajuus ei kasva lineaarisesti, vaan kahden vierekkäisen taitotason välillä on jossakin kohtaa asteikkoa (esim. siirryttäessä tasolta A2 tasolle B1) muita suurempi hyppäys. Aiemmassa tutkimuksessa on käynyt ilmi (Martin ym. 2010), että virheiden määrä usein vastaavasti alenee merkittävästi kahden seuraavan taitotason välillä (esim. siirryttäessä tasolta B1 tasolle B2).

Tässä tutkimuksessa *A*- ja *U*-verbin valintavirheitä on etenkin yleisluontoisissa kirjoitelmissa niin koululaisilla kuin aikuisillakin koko aineistoissa vain muutamia. Virheiden määrän väheneminen ei siis toimi taitotason indikaattorina näissä teksteissä. Sen sijaan akateemisissa tekstilajeissa virheitä on huomattavasti enemmän. Näyttääkin siltä, että tämä *U*-verbien käytön virheiden suhteellinen runsaus onkin osoitus korkeasta taitotasosta ja vaativasta tekstilajista. Tietenkin myös kirjoitustilanne voi vaikuttaa: YKI-aineisto on kirjoitettu kielikoetilanteessa, LAS2-aineistossa taas tenttivastauksetkin on kirjoitettu osoittamaan sisällöllistä tietämystä, ei kielitaitoa. Vaikka YKI-kokeen arviointikriteerit korostavat kielenkäyttötaitoja, useimmat osallistujat ovat varmasti tottuneet varomaan kielivirheitä koesuorituksissa eivätkä ole halukkaita ottamaan riskejä, vaan valitsevat sellaisia sanoja ja rakenteita, jotka arvelevat hallitsevansa hyvin.

Toinen syy, joka saattaa vaikuttaa *U*-verbivirheiden määrään LAS2-aineistossa, on se, että kirjoittajat ovat suomen kielen opiskelijoita ja heillä on teoreettista tietoa transitiivisten *A*-verbien ja intransitiivisten *U*-verbien eroista. Siinä missä koululaiset ja useimmat YKI-osallistujat kirjoittavat suomea pitkälti ”korvakuulolta”, LAS2-aineiston kirjoittajat prosessoivat kieltä tietoisemmin. Tämä näkyy erityisesti uudis-

muodosteista ja muutamista muista käyttövirheistä, jotka voi ymmärtää kielioppien *U*-verbien käyttöä koskevien sääntöjen pohjalta (mm. esimerkki (37) luvussa 6 ja esimerkit (70) ja (71) luvussa 7). *U*-verbikonstruktioiden, kuten kaikkien konstruktioiden, käytön rajoitukset ovat kuitenkin pitkälti sanakohtaisia eivätkä opittavissa pelkästään sääntöjen avulla.

Koska LAS2-aineistoa ei ole teksti tekstiltä arvioitu, ei tässä ole mahdollista päätellä, olisiko *U*-verbivirheiden väheneminen mahdollisesti taitotasoja C1 ja C2 erottava tekijä. Arviointi olisi syytä tehdä, sillä näitä ylimpiä taitotasoja erottavia lingvistisiä ilmiöitä ei ole toistaiseksi tutkimuksessa löydetty.

Kolmas tapa etsiä funktionaalisten taitotasojen ja lingvististen ilmiöiden suhdetta on tarkastella, millaisissa konteksteissa ilmiö esiintyy eri taitotasoilla. Verbivalikoima tietenkin lisääntyy ylemmillä taitotasoilla, kuten sanavarasto ylipäättään. Samojakin verbejä käytetään kuitenkin myös eri tavoin alkeistasolla ja edistyneiden oppijoiden teksteissä. Kysymys näyttää olevan ainakin kahdesta eri ilmiöstä: verbien käyttökehys monipuolistuu, ja verbit ovat mutkikkaampien rakenteiden osia. Kyse on siis pitempien ja rakenteellisesti kompleksisempien jaksosten hallinnan kehittymisestä (vrt. Reiman 2011). Toisaalta kyse on myös abstraktiotason noususta. Osittain tämä on seurausta aineiston aihepiireistä, mutta voi myös kysyä, olisiko abstraktisuus myös eräs kehittyneen kielitaidon komponentti sujuvuuden, tarkkuuden ja kompleksisuuden rinnalla (ks. Martin ym. 2010). Tämän ehdotuksen todentaminen vaatisi kuitenkin toisentyypistä aineistoa, jossa kirjoittajat olisivat abstraktisuustasoltaan samanlaisen ja samaa aihepiiriä käsittelevän tehtävän edessä.

Lähteet

- Alanen, Riikka, Ari Huhta, Mirja Tarnanen 2010. Developing and assessing L2 writing tasks across CEFR proficiency levels. – Inge Bartning, Maisa Martin, Ineke Vedder (Eds.). *Communicative Proficiency and Linguistic Development, Intersections Between SLA and Language Testing Research*. EUROSLA Monograph Series 1. European Second Language Association, 21–56. <http://eurosla.org/monographs/EM01/EM01home.html> (9.3.2012).
- Cefling. <https://www.jyu.fi/hum/laitokset/kielet/cefiling> (24.08.2012).
- Denisyuk, Anna 2010. Sanavalintavirheiden tyypit ja taustaa suomenoppijoilla. Julkaisematon pro gradu -tutkielma. Turun yliopisto.
- Ellis, Rod 2008. *The Study of Second Language Acquisition*. Second Edition. Oxford: Oxford University Press.
- EVK = Eurooppalainen viitekehys. Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys 2004. Helsinki: WSOY.
- Glahn, Ester, Gisela Håkansson, Björn Hammarberg, Anne Holmen, Anne Hvenekilde, Karen Lund 2001. Processability in Scandinavian second language acquisition. – *Studies in Second Language Acquisition* 23 (3), 389–416. <http://dx.doi.org/10.1017/S0272263101003047>
- Goldberg, Adele 1995. *Constructions. A construction grammar approach to argument structure*. Cognitive theory of language and culture. Chicago, IL: The University of Chicago Press.
- Hawkins, John A., Paula Buttery 2010. Critical features in learner corpora: Theory and illustrations. – *English Profile Journal* 1 (1), e5. <http://dx.doi.org/10.1017/S2041536210000103>
- Hopper, Paul J., Sandra A. Thompson 1980. Transitivity in grammar and discourse. – *Language* 56 (2), 251–299. <http://dx.doi.org/10.1353/lan.1980.0017>
- ICLFI. Kansainvälinen suomi vieraana kielenä -korpus, Oulun yliopisto. <http://www oulu.fi/hutk/sutvi/oppijankieli/index.html> (17.08.2012).
- Koivisto, Vesa 1991. Suomen verbikantaisten *UtU*-verbijohdosten semantiikkaa. *Suomi* 161. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Koivisto, Vesa 1995. Itämerensuomen refleksiivit. *Suomalaisen Kirjallisuuden Seuran toimituksia* 622. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Kulonen-Korhonen, Ulla 1985. Deverbaalisten *U*-johdosten semantiikkaa. – *Viritäjä* 89 (2), 290–309.
- OPS = Opetussuunnitelmien ja tutkintojen perusteet. Verkkolähde: http://www.oph.fi/saadokset_ ja_ohjeet/opetussuunnitelmien_ ja_ tutkintojen_perusteet (9.3.2012).

- Martin, Maisa 2000. *U*-verbit – suomalaista hienotunteisuutta parhaimmillaan? – *Virittäjä* 104 (2), 289–295.
- Martin, Maisa 2008. Negation as a marker of proficiency level in L2 Finnish. – Suullinen esitys Language Learning -lehden sponsoroimassa Round Table -tilaisuudessa. EUROSLA 18. Aix-en-Provence, Ranska, 10.–13.9.2008.
- Martin, Maisa tulossa. Sentences and clauses as complexity measures in second language writing: a segmentation experiment. – Julkaistaan Greifswaldin yliopiston suomen kielen opetuksen juhlakirjassa.
- 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 Monograph Series 1. European Second Language Association, 57–80. <http://eurosla.org/monographs/EM01/EM01home.html> (9.3.2012).
- Pallotti, Gabriele 2007. An operational definition of the emergence criterion. – *Applied Linguistics* 28 (3), 361–382. <http://dx.doi.org/10.1093/applin/amm018>
- Pienemann, Manfred 1998. *Language Processing and Second Language Development. Processability Theory*. *Studies in Bilingualism* 15. Amsterdam: John Benjamins.
- Puro, Tarja 2002. Suomi toisena kielenä – aikuisoppijan verbien kehittyminen alkeiskurssilla. Julkaisematon lisensiaatintutkielma. Jyväskylän yliopisto. <https://jyx.jyu.fi/dspace/handle/123456789/11904> (7.3.2012).
- Reiman, Nina 2011. Two faces of complexity: structural measures and diversity of constructions. – *Nordand* 6 (2), 9–23.
- Siitonen, Kirsti 1993. Intransitiivinen poikkeama systeemistä. – Sirkka Saarinen, Jorma Luutonen, Eeva Herrala (Toim.). *Systeemi ja poikkeama*. Juhlakirja Alho Alhoniemen 60-vuotispäiväksi 14.5.1993. Suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 42. Turku: Turun yliopisto, 312–320.
- Siitonen, Kirsti 1999. Agenttia etsimässä. *U*-verbijohdokset edistyneen suomenoppijan ongelmana. Suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 63. Turku: Turun yliopisto.
- Siitonen, Kirsti 2002. Intentionaalisuusasteen tunnistaminen suomenkielisestä ilmauksesta. – Anna Mauranen, Liisa Tiittula (Toim.). *Kieli yhteiskunnassa – yhteiskunta kielessä*. AFinLA:n vuosikirja 60. Jyväskylä: Suomen soveltavan kielitieteen yhdistys, 287–304.
- Tomasello, Michael 2003. *Constructing a Language. A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

YKI = Yleiset kielitutkinnot. Verkkolähde: http://www.oph.fi/yleiset_kielitutkinnot (9.3.2012).

Verspoor, Marjolijn, Kees de Bot, Wander Lowie 2011. A Dynamic Approach to Second Language Development: Methods and Techniques. Amsterdam: John Benjamins.

Wolfe-Quintero, Kate, Shunji Inagaki, Kim Hae-Young 1998. Second language development in writing: measures of fluency, accuracy, and complexity. Technical Report No. 17. University of Hawai'i at Mānoa: Second Language Teaching & Curriculum Center.

Kirsti Siitonen

Turun yliopisto, suomen kieli, kieli- ja käännöstieteiden laitos
20014 Turku, Finland
kisiito@utu.fi

Maisa Martin

Jyväskylän yliopisto, kielten laitos
PL 35 (F)
40014 Jyväskylä, Finland
maisa.martin@jyu.fi

Liite 1. Viitekehyksen taitotasot

Peruskielitaito:	tasot A1 ja A2
Itsenäisen kielenkäyttäjän kielitaito:	tasot B1 ja B2
Taitavan kielenkäyttäjän kielitaito:	tasot C1 ja C2

Liite 2. Aineistosta haetut verbit

aiheuttaa	liittyä	paistaa	säilyä
aiheutua	löytää	paistua	särkeä
heijastaa	löytyä	parantaa	särkyä
heijastua	maistaa	parantua	säästää
jatkaa	maistua	peittää	säästyä
jatkuu	muistaa	peittyä	toistaa
johtaa	muistua	pelastaa	toistua
johtua	muodostaa	pelastua	tuntea
järjestää	muodostua	pelästyttää	tuntua
järjestyä	muuttaa	pelästyä	unohtaa
kaataa	muuttua	perustaa	unohtua
kaatua	myöntää	perustua	vahvistaa
kehittää	myöntyä	peruuttaa	vahvistua
kehittyä	naarmuttaa	peruuntua	vaihtaa
kerätä	naarmuttua	sairastaa	vaihtua
kertyä	nähdä	sairastua	valmistaa
kiinnostaa	näkyä	siirtää	valmistua
kiinnostua	näyttää	siirtyä	väsyttää
kuulla	näyttäytyä	sytyttää	väsyä
kuulua	painaa	syttyä	yhdistää
liittää	painua	säilyttää	yhdistyä

***U*-stem verbs as indicators of proficiency level**

KIRSTI SIITONEN

Turun yliopisto

MAISA MARTIN

Jyväskylän yliopisto

The Finnish lexicon contains a group of verbs which are derived from other verbs with the derivational suffix *U*, such as *vaihtaa* > *vaihtua* ‘to change, exchange’ or *jatkaa* > *jatkua* ‘to continue’. The *A*-stemmed verbs are generally transitive, while their *U*-stemmed counterparts are often reflexive or automative, indicating that something happens unintentionally or without a human agent, as opposed to the Finnish passive forms which imply an unknown human agent. In practice, however, the semantic and syntactic relationships between these two groups of verbs are not far from clear, varying from one verb to another, which makes even the most advanced learners of Finnish struggle with these verbs.

In this paper three sets of written learner data are examined to determine whether the use of *U*-verbs could be seen as a determinant of proficiency level. School data, written by 12–16 year old second language (L2) users of Finnish, represents the functional Common European Framework of Reference (CEFR) levels from A1–B2. The National Certificates (YKI) data is from adults writing at levels A1–C2. The data from Turku University foreign students of Finnish and related languages (LAS2) represents academic L2 writing at levels B2–C2. *A/U* pairs of verbs were collected from this data (totalling about 182,000 words) according to the list in Appendix 2.

The results show that the occurrence of the verbs in the list vary from 5,6 (YKI–A1) to 17,4 (examination answers of the students) per 1000 words. The variety and use of these verbs clearly grows as proficiency develops. Furthermore, the advanced students use more of the *U*-verbs and less of the *A*-verbs. Thus the rich variety of *U*-verbs can be seen as an indicator of high level.

The accuracy of the use of these verbs was also studied. Errors pertaining to the syntacto-semantic use of these verbs were rare among the school children and even the YKI adult group. It seems that the few *U*-verbs they used had been learnt as part of ready-made constructions rather than as individual verbs in contrast with their derivative pairs. There were more errors in the writing of the

advanced students who were more likely to process their verb choices using the semantic and syntactic knowledge they had learned during their studies. They also produced innovative forms, filling the gaps which exist in the derivational paradigms.

The third dimension studied was the sentence contexts where the verbs were used. As can be expected, the contexts in which the more advanced learners used the verbs were syntactically complex and contained longer sequences. The topics were more abstract and the style academic, which often made the writers overextend themselves and lose control over this particular syntacto-semantic device.

The *U*-verbs can thus be seen as indicators of proficiency in two ways: the variety and overall occurrence of *U*-verbs increases steadily as the proficiency level grows, particularly in academic genres. On the other hand, the number of errors grows as well, leading into an unusual situation where the presence of errors, not their absence, is a sign of high proficiency. This is due to the characteristics of Finnish academic and literary style where these verbs are very common and used in complex constructions, making the processing burden overwhelming even for the most advanced students. Another reason is that the syntacto-semantic differences between the Finnish passive and the *U*-verbs are very fine-grained and involve expressing semantic and pragmatic notions, such as intentionality, which in many languages have not grammaticalized expression.

Keywords: accuracy; Finnish *U*-verbs; indicators of language proficiency; language proficiency levels