

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Martin, Maisa

**Title:** Käyttötaajuus ja tarkkuus toisen kielen kehityksessä

**Year:** 2022

**Version:** Published version

**Copyright:** © 2022 Suomen soveltavan kielitieteen yhdistys AFinLA

**Rights:** In Copyright

**Rights url:** <http://rightsstatements.org/page/InC/1.0/?language=en>

**Please cite the original version:**

Martin, M. (2022). Käyttötaajuus ja tarkkuus toisen kielen kehityksessä. In L. Aarikka, K. Priiki, & I. Ivaska (Eds.), *Soveltavan kielitieteilijän sormenjälkiä etsimässä : kielen rakenteet ja niiden käyttäjät* Kirsti Siitosen tutkimuksellisia kiintopisteinä (pp. 81-102). Suomen soveltavan kielitieteen yhdistys AFinLA. AFinLA-teema. <https://doi.org/10.30660/afinla.111254>

# OSIO II

**Rakenteet käytössä ja oppimisessa**

Aarikka, L., K. Priiki & I. Ivaska (toim.) 2022. Soveltavan kielitieteilijän sormenjälkiä etsimässä: Kielen rakenteet ja niiden käyttäjät Kirsti Siitosen tutkimuksellisinä kiintopisteinä. *In search of the fingerprints of an applied linguist: Linguistic structures and their users as scholarly focal points in Kirsti Siitonen's career. AFinLA-teema / n:o 14. 81–102.*

**Maisa Martin**  
Jyväskylän yliopisto

## Käyttötaajuus ja tarkkuus toisen kielen kehityksessä

The path of learning a new language must find a route between the need to express oneself and the desire to do so accurately. The article discusses this relationship between learning to use various constructions and getting them right. The pseudolongitudinal data consist of about 1200 texts written by young and adult learners of Finnish as a second language, assessed to represent the functionally determined Common European Framework of Reference (CEFR) levels A1–C2. The results are sifted from a large number of studies which have followed the use and accuracy of a given construction in these texts. The results show that an increase of use often precedes an improvement in accuracy at the next CEFR level. Common constructions peak earlier both in use and accuracy than the less obligatory ones but errors in rarer constructions can function as indicators of high-level language skills.

**Keywords:** frequency of use, accuracy, second language development,  
Finnish as a second language

**Asiasanat:** taajuus, tarkkuus, toisen kielen kehitys, suomi toisena kielenä

## 1 Johdanto

Kielenoppimisen kulku polveilee ilmaisutarpeiden ja kohdekielenomaisuuteen pyrkimisen välimaastossa. Uuden kielen käyttäjällä on sanottavaa, mutta ei vielä käytössään kielellisiä keinoja, joita asian ilmaiseminen vaatisi. Silloinkin kun asia käy selväksi, ilmaus ei aina kaikilta osin ole samanlainen kuin kohdekielessä on tapana. Tässä luvussa tarkastelen joidenkin ilmausten ja rakenteiden (konstruktioiden) esiintymistaajuuden ja tarkkuuden kehitystä suomea toisena kielenä kirjoittavien teksteissä Eurooppalaisen viitekehyksen (2003, EVK) taitotasoilla, jotka on määritelty funktionaalisesti. Kyseessä on katsaus, johon on koottu monien suunnilleen samanlaiseen aineistoon perustuvien tutkimusten tuloksia. Tavoitteena on pohtia, kumpi oli ensin, muna vai kana, eli kehittykö tarkkuus vasta käytön myötä vai onko ainakin jonkinasteinen tarkkuus konstruktion käytön edellytys. Laajempaa viitekehyksenä on pyrkimys etsiä funktionaalisten taitotasojen mahdollisia kielellisiä indikaattoreita, koska niiden tunnistamisella on merkitystä niin taitotasoajattelun kuin kielenopetuksenkin kannalta. Käyttötaajuuden ja tarkkuuden suhteen muutokset voisivat siis olla yksi tällainen indikaattori.

Kielellisen konstruktion tarkkuuden ja käytön välinen suhde sisältää myös sen, että jotain konstruktiota ei käytetä, vaikka sille olisi tarvetta – tarpeettomia konstruktioita on vaikea kuvitella. Tämä kysymys nousee esille Kirsti Siitosen tutkimien U-verbien (esim. *jatkua, vaihtua* vs. *jatkaa, vaihtaa*) osalta. Ensikielenään suomea puhuvat käyttävät U-verbijä luontevasti aivan arkisissakin yhteyksissä, mutta ne ilmaantuvat oppijoiden käyttöön vasta korkeimmilla taitotasoilla ja niissä esiintyvät epätarkkuudet näyttävät pikemminkin todisteina oppimisesta kuin kielitaidon heikkouksista. Käytön ja tarkkuuden suhde voi siis olla monenlainen.

Käytän pohdiskelun pohjana Jyväskylän yliopiston Cefling- ja Topling-hankkeiden (ks. luku 3) piirissä tehtyjä tutkimuksia, joiden aineistona on aikuisten ja nuorten kirjoittamia, kolmea tekstilajia edustavia tekstejä. Nämä on arvioitu EVK:n taitotasoille A1–C2, joten aineisto tarjoaa näennäispitkittäisen näkökulman kirjoitetun kielen oppimiseen. Mukana on myös tekstejä Turun yliopiston LAS2-korpuksesta. Kyse ei siis ole yksittäisen oppijan kehityksestä, vaan yleiskuvasta. Aineistosta on tehty lukuisia väitöskirjoja ja maisterintutkielmia, joiden tuloksia hyödynnän katsauksessa.

Esittelen aluksi hyvin lyhyesti muutamia käsitteitä ja työn lähtökohtia. Seuraavassa luvussa kuvataan ne aineistot, jotka ovat tämän luvun esittelemien tutkimusten taustalla. Tuloslukuun on taulukoitu kunkin tutkimuksen käsittelemän ilmiön käyttötaajuuden ja tarkkuuden kehitys taitotasolta toiselle. Tutkimukset on jaettu neljään alaryhmään. Viimeisessä luvussa arvioin tulosten merkitystä ja pyrin vastaamaan edellä esittämäni kysymykseen käyttötaajuuden ja tarkkuuden suhteesta.

Käyttötaajuuden ja tarkkuuden kehityksen suhdetta tarkastellaan tässä luvussa hyvin väljästi ja yleisellä tasolla. Useimmissa taustalla olevissa tutkimuksissa ei ole käytetty tilastollista analyysiä esimerkiksi havaittujen erojen merkitsevyydestä. Koska kyseessä on näennäispitkittäistutkimus, mukana ei ole sellaisia tavallisia muuttujia kuin oppijan ensikieli tai koulutustausta. Aineisto antaisi mahdollisuuden tekstilajikohtaisten erojen analyysiin ja sitä onkin useimmissa tässä esitellyissä tutkimuksissa tehty, mutta tämä näkökulma ei mahdu mukaan. Tarkasteltavien tutkimusten lukuisuuden vuoksi tässä ei ole mahdollista esitellä kunkin tutkimuksen käsitteellisiä tai aineistoon liittyviä rajanvetoja, jotka ovat kuitenkin löydettävissä alkuperäistutkimuksista. Tavoittelen siis ennen kaikkea laaja-alaista katsausta siihen, millaisia ilmiöitä oppijoiden kirjoituksessa eri taitotasolla esiintyy.

## 2 Käsitteitä ja lähtökohtia

### 2.1 CAF-triadi ja DEMfad-malli

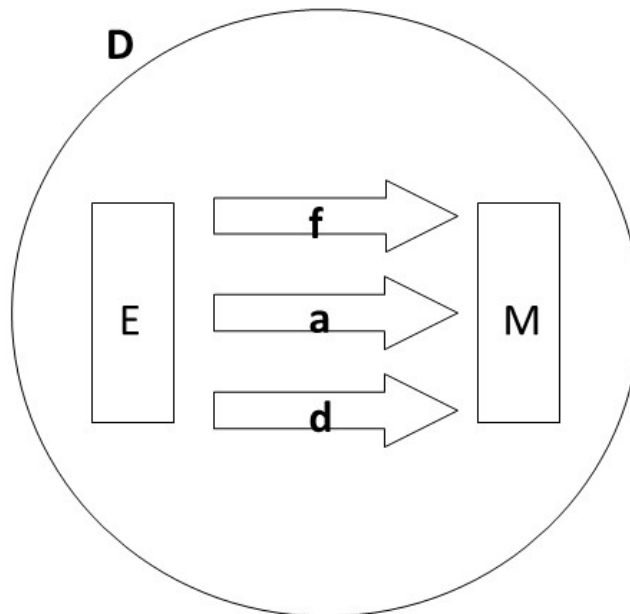
Toisen kielen oppimisen tutkijat puhuvat usein CAF-triadista (Complexity, Accuracy, Fluency) eli kielen kompleksisuuden, tarkkuuden ja sujuvuuden suhteesta. Näiden kolmen muuttujan kannalta on tarkasteltu oppijankielen kehitystä 1990-luvun lopulta lähtien (esim. Skehan 1998, 2009; Housen ym. 2012). Vaikka mitään näistä kolmesta termistä ei ole onnistuttu aukottomasti tai kiistattomasti määrittelemään, ne näyttävät kuitenkin olevan intuitiivisesti riittävän selkeitä toimiakseen kielen kehityksen tutkimuksen ja kielenopetuksen apuvälineinä.

Cefling- ja Topling-hankkeissa pyrittiin useita eri konstruktioita tutkimalla luomaan yleiskuvaa kielellisten piirteiden kehityksestä suhteessa funktionaalisesti määriteltyihin taitotasoihin. Erillistutkimusten vertailtavuuden vuoksi tarvittiin yhteinen malli, jonka avulla kehitystä seurataan. Tähän tarkoitukseen kehitettiin CAF-ajattelun pohjalta DEMfad-malli (Kuvio 1; Franceschina ym. 2006; Martin ym. 2010; Kaivapalu ym. 2018), jota on sittemmin käytetty useissa muitakin aineistoja koskevissa tutkimuksissa (esim. Ivaska 2015, Kitsnik 2018, Siivelt & Mustonen 2013). CAF-ajattelun mukaista kolmijakoa kehityksen parametreihin muokattiin kirjoitettuun suomenkieliseen aineistoon paremmin sopiviksi. Tarkkuus säilytettiin sellaisenaan, mutta sujuvuus korvattiin käyttötaajuudella (ks. 2.3).

Kolmas muuttuja nimettiin kompleksisuuden asemasta distribuutioksi ja sen määrittely jätettiin mallia suunniteltaessa tarkoituksellisesti avoimeksi, koska CAF-triadin pohjalta tehtyjen tutkimusten käyttämät kompleksisuuden määritelmät ja mittarit eivät näyttäneet riittävän informatiivisilta. Ajatuksena oli, että kukin tutkija voisi tarkastella kolmatta dimensiota tutkimuskohteensa näkökulmasta parhaaksi katsomallaan tavalla ja niin onkin käynyt. Näin syntymässä olevan uuden "komplek-

sisuuden” määritelmän koonti on kuitenkin vielä kesken, joten jätän kompleksisuuden kehityksen tässä tarkastelun ulkopuolelle, sillä siitä tiedetään suomen kielen osalta vasta liian vähän (ks. kuitenkin Mylläri 2020a, 2020b ja Reiman 2011b). Muihin kieliin, etenkin Englantiin, käytetyt kompleksisuuden määrälliset mittarit eivät suomi toisena kielenä -aineistossa näytä juuri erottavan taitotasoa toisistaan (Mylläri 2020b).

DEMfad-mallissa D(omain) viittaa tutkittavaan konstruktion, E(mergence) taitotasoon, jolloin se ensimmäisen kerran esiintyy tutkittavassa aineistossa, ja M(astery) taitotasoon, jolloin se hallitaan (alaluku 2.2).



KUVIO 1. DEMfad-malli

Mallissa seurataan taitotasolta toiselle käyttötaajuuden (frequency), tarkkuuden (accuracy) ja distribuution kehitystä. Tarkkuuden osalta voidaan määrittellä numeerinen tavoitetaso (Mastery, alaluku 2.2), käyttötaajuuden ja distribuution osalta tavoitteen määrittely olisi mahdollista vain hyvin suurten ja tekstilajiltaan yhtenäisten tekstimassojen avulla ja pelkästään probabilistisesti. Yksilöllinen kielenkäyttöhan on aina individuaalista ja dynaamista, jatkuvasti muuttuvaa. Tässä luvussa tarkastellaan kuitenkin käyttöä ja tarkkuutta kullakin funktionaalisella taitotasolla ottamatta kantaa siihen, saavutetaanko jokin tavoitetaso.

DEMfad-mallin juuret ovat siis CAF-ajattelussa, sen kolmijako nähdään arvokkaana mutta ei täydellisenä lähtökohtana kielenkehityksen tarkastelulle. DEMfadin

teoreettisena etätavoitteena on siten ottaa kantaa CAF-triadin yleisiin oletuksiin ja mahdollisiin ongelmakohtiin.

## 2.2 Tarkkuus

Kielenoppijoiden kielestä etsitään ja löydetään *virheitä*. Niiden ajatellaan erottavan oppijan kielen äidinkielen puhujan kielestä. Niin *oppija* kuin *äidinkielen puhuja*kin ovat kuitenkin abstraktioita, joita on mahdoton määritellä tarkasti. Milloin ihminen lakkaa olemasta kielenoppija? Kuka on se äidinkielen puhuja, johon verrataan? (Näitä kysymyksiä käsittelee esimerkiksi Ruuska 2020; ks. myös Suni & Ruuska tässä julkaisussa). Määrittelyn ongelmallisuudesta huolimatta virheen käsitettä tarvitaan, kun tarkastellaan kielenoppimisen kulkua. Virheet eivät ole vain jotain, joka pitäisi kitkeä pois, vaan myös oppimisen edellytys ja indikaattori. Tästä syystä niiden määrää ja laatua tarkastellaan, kun halutaan tietää, miten kielenoppiminen etenee. Käsite *tarkkuus* on sävyllään myönteisempi kuin virhe ja vakiintunut käyttöön CAF-ajattelun myötä. Muita usein käytettyjä termejä ovat *kohdekielenomaisuus* ja *konventionaalisuus* (ks. esim. Schmid 2020).

Olipa ilmiön nimi mikä tahansa, jokaisessa tutkimuksessa on määriteltävä erikseen, mikä milloinkin katsotaan virheeksi. Cefling-hankkeen osatutkimuksissa on tarkasteltu eri konstruktoiden kehitystä taitotasoarvioidussa aineistossa. Tällöin tarkkuuden kriteerinä on kyseisen konstruktion oikeellisuus kontekstissaan. Jos siis tutkimuksen kohteena on muoto, esimerkiksi kielto eri aikamuodoissa, oikeaksi ajatellaan ilmaus, jonka voi tunnistaa kyseisen aikamuodon kieltomuodoksi, vaikka oikeinkirjoitus ei ihan onnistuisikaan. Esimerkiksi *ei vastanut* tai *en ossa* eivät aikamuodon tunnistamisen kannalta ole virheellisiä, mutta *en söi* on. Vaikka se on tunnistettavissa kieltomuodoksi, ei voi tietää, tavoitellaanko preesens- vai imperfektimuotoa. Jos taas tutkimuksen kohteena on vaikkapa paikallissijojen käyttö, on samantekevää, onko illatiivi muodostettu oikein, jos se on tunnistettavissa illatiiviksi ja käyttökonteksti on oikea. Tämän artikkelin pohjana olevissa tutkimuksissa tarkkuus siis määrittyy eri tavoin sen mukaan, mihin tutkimus kohdistuu, mutta jossain suhteessa virheeksi luokiteltu ilmaus aina poikkeaa kohdekielessä tavanomaisesta.

Tarkkuuden osalta tässä esiteltyjen tutkimusten tuloksia verrataan Brunnin ym. (2019) artikkelin aikuisten suomea vieraana kielenä oppivien aineistosta saatuihin tuloksiin (ks. luku 4.6). Kyseisen aineiston virheet on annotoitu ottamatta huomioon, mitä kulloinkin ollaan tutkimassa. Tämänäyttöinen virheen määrittely on tietysti yleispätevämpi, vaikka siinäkin on tuki ongelmia. Se ei esimerkiksi ota huomioon tekstilajia tai tehtävänantoa. Usein on myös vaikea määritellä, mihin virhekategoriiaan kukin tapaus kuuluu (ks. Brunni ym. 2015, Jantunen 2011, 2014). Luokittelun ja rajanveto-ongelmat ovat ikuisia, mutta toisaalta kohdekielen puhujilla on usein melko yhtenäinen intuitiivinen käsitys siitä, mikä heidän kielessään on mahdollista ja mikä ei. Silläkin tiedolla on arvoa oppijalle, joka pyrkii parantamaan kielitaitoaan.

DEMfad-malliin perustuvissa tutkimuksissa tarkkuuden hallinnan (Mastery, ks. alaluku 2.1) tasoksi on asetettu 80%. Tämä pohjautuu Hyltenstamin ja Abrahamssonin (2003) tutkimukseen erittäin taitavien toisen kielen käyttäjien ja äidinkielisten puhujien taitojen suhteesta. Kaikki heidän tutkimansa äidinkielisten pääsivät tarkkuudessa 80 %:n tasolle, mutta kaikki eivät 100 %:iin. Oppijankielen tutkimuksissa on joskus käytetty myös 50 %:n tarkkuutta, koska se kertoo kuitenkin jo kyseisen ilmiön melko hyvästä hallinnasta (esim. Glahn ym. 2001). Toisaalta yhtäkin virheetöntä esiintymää voidaan pitää merkinä säännön oppimisesta (esim. Pienemann 1998), jos voidaan näyttää, ettei kyse ole ulkoa opitusta fraasista. Siinä ei kuitenkaan oikeastaan ole kyse tarkkuudesta vaan jonkin rakenteen ilmaantumisesta oppijan kieleen. Konstruktionäkökulmasta (luku 2.5) kielenoppimista tarkastelevat taas lähtevät siitä, että jonkin konstruktion esiintyminen ensimmäisen kerran osoittaa jo sen oppimista (esim. Martin ym. 2010), vaikka ilmaus ei olisikaan kohdekielen mukainen.

### 2.3 Käyttötaajuus

Kertaus on opintojen äiti eli toiston määrä vaikuttaa oppimisen kulkuun. Tämä on yksi käyttöpohjaisen kielenoppimisenäkemyksen kulmakivistä ja sen vaikutusta on tutkittu melko paljon (ks. esim. Ellis 2016). Usein käytetyt sanat, rakenteet ja ilmaukset juurtuvat muistiin ja löytyvät sieltä helpommin kuin harvinaiset. Siten kielenkäyttötilanteissa usein nähty ja kuultu tulee todennäköisemmin omaankin käyttöön. Tästä juurtumisprosessista käytetään myös termiä *entrenchment* (Schmid 2020).

Cefling- ja Topling-hankkeissa on selvitetty tarkasteltavina olevien konstruktioiden käyttötaajuutta yhtenä kehityksen indikaattorina. Käyttötaajuus on etäistä sukua sujuvuudelle, joka on yksi CAF-triadin parametreista, vaatiihan taajuuden kasvu taustakseen jonkinlaista tuottamisen lisääntymistä, mutta kirjoitettuja tuotoksia tutkimalla ei sujuvuuteen pääse käsiksi. Toki kirjoittamisen sujuvuuttakin voidaan tutkia ja Cefling- ja Topling-hankkeiden piirissä on muiden kuin suomen kielen osalta niin tehtykin, mutta tämän artikkelin pohjana ovat vain valmiita tuotoksia koskevat tutkimukset. Hankkeita suunniteltaessa haluttiin ottaa mukaan näkökulma, joka kuvaa tuottamisen määrää monipuolisemmin kuin tavallisemmin käytetty sanojen tai lauseiden määrä, ja tällaiseksi valikoitui tutkittavana olevan konstruktion käyttötaajuus tuhatta sanaa kohti. Näin konstruktion käyttöä voidaan verrata taitotasojen välillä tekstimassan kokonaissanamäärästä riippumatta ja tarvittaessa myös verrata eri konstruktioiden esiintymistä aineistossa.

### 2.4 Käyttötaajuuden ja tarkkuuden suhde

Jonkin kielellisen ilmiön käyttötaajuuden ja tarkkuuden suhdetta voi lähestyä monelta suunnalta. Tässä on valittu hyvin yleinen, taitotasoihin jaettuun aineisto-



massaan pohjautuva näkökulma, jonka päämääränä on tavoittaa jotakin säännönmukaisuutta funktionaalisten taitotasojen ja kielellisten taitojen suhteesta. Toista ääripäätä, yksilön kielen yksityiskohtaisen kehityksen seuraamista, edustaa Complex Dynamic Systems Theory (CDST), jonka keskeisiä tutkimuskohteita myös on opittavina olevien ilmiöiden suhde toisiinsa. Uusi ilmiö voi joko kilpailla aiemmin opitun kanssa, jolloin aiemmin opittu heikkenee taajuuden tai tarkkuuden tai molempien suhteen. Uusi ja aiemmin opittu voivat myös tukea toisiaan. Kolmas vaihtoehto on ehdollinen suhde: jokin ilmiö on opittava ennen kuin toisen oppiminen mahdollistuu. (Spoelman & van Dijk 2011.) Suomi toisena kielenä -alalla tätä menetelmää edustavat Spoelman & Verspoor (2010), Tilma (2014) ja Lesonen (2020). Koko kielisysteemin synnyn ja oppimisen näkökulmasta taas samoja ilmiöitä eli konventionaalistumisen ja vakiintumisen (entrenchment) suhdetta tarkastelee laajasti Schmid 2020.

## 2.5 Konstruktio

Cefling- ja Topling-hankkeiden teoreettinen tausta on käyttöpohjaisessa teoriassa kielenoppimisesta (ks. esim. Bybee & Beckner 2015). Sen mukaan kieli opitaan käyttämällä sitä ja abstrahoiden kielenkäytöstä konstruktioita, joita voidaan yhdistää, laajentaa ja muokata (ks. esim. Goldberg 1995, Tomasello 2003). Termillä konstruktio viitataan kielen yksikköön, jolla on fonologinen tai ortografinen, morfologinen ja syntaktinen muoto, semanttinen merkitys ja pragmaattinen käyttö vuorovaikutustilanteessa. Lähtökohta on sama kuin konstruktiokieliopissa, mutta hankkeissa ei ole sitouduttu mihinkään tiettyyn teoreettiseen käsitykseen tai notaatioon, vaan konstruktion käsitettä on tulkittu laajasti. Jokaisessa osatutkimuksessa on erikseen määritelty, mitä katsotaan kuuluvaksi tutkittavan konstruktion piiriin. Useimmissa tässä esitellyissä tutkimuksissa konstruktioita on lähestytty sen rakenteen kannalta (alaluvut 4.2 ja 4.3), mutta myös merkitys ja käyttö toimivat tulokulmina (alaluku 4.4). Myös sanaa voidaan pitää konstruktiona, joten mukaan on otettu myös pari sanastotutkimusta (alaluku 4.5). Kuvauksia yksilötason abstrahoitumisprosesseista kielenoppimisen edistyessä ovat esimerkiksi Eskildsen (2008) tai Lesonen (2020). Eri taitotasojen tuotoksissa näkyviä abstrahoitumisprosesseja on löydettävissä lähes kaikkien luvussa 4 esiteltujen tutkimusten esimerkeistä.

## 3 Aineisto ja menetelmä

The Linguistic Basis of the Common European Framework levels (Cefling, 2007–2009) ja Toisen kielen oppimisen polut (Topling, 2010–2013) olivat Suomen akatemian rahoittamia tutkimushankkeita. Niiden päätavoitteena oli selvittää, onko löydettävissä kielellisiä piirteitä, joiden esiintyminen ja kehittyminen ovat yhteydessä funktionaa-

lisesti määriteltyn kielitaidon tasoon. Näin on pyritty saamaan tietoa oppijankielen kehityksen yleisistä linjoista mutta myös keräämään tietoa opetuksen ja kielitaidon arvioinnin tueksi. Tutkittavia kieliä ovat suomi toisena kielenä (S2) ja ruotsi ja englanti Suomessa. Tässä luvussa käsitellään vain S2-tutkimuksia.

Hankkeiden kokoamissa aineistoissa on eri tekstilajeja edustavia kirjoitelmia eri-ikäisiltä koululaisilta ja opiskelijoilta. Cefling-hankkeen S2-perusaineistossa on 527 tekstiä ja 26 000 sanetta. Nämä yläkoululaisten tekstit edustavat taitotasoa A1–B2, vaikka B2-tason tekstejä on muita vähemmän eikä niitä ole kaikissa tutkimuksessa otettu mukaan. Topling-aineistossa on 2548 tekstiä ja 101 000 sanetta ja ne edustavat taitotasoa A1–C2, sillä mukana on myös lukioikäisiä nuoria. Näitä aineistoja koskevat tiedot ovat tässä nimellä ”nuoret”. Topling-aineistossa on myös kirjoitelmia samoilta kirjoittajilta kolmena peräkkäisenä vuotena, joten sen avulla on mahdollista tehdä myös pitkittäistutkimusta. Sitä ei kuitenkaan toistaiseksi ole tehty. Valtaosa tässä esitellyistä nuorten kirjoittamista koskevista tiedoista perustuu Cefling-aineistoon.

Hankkeissa käytetty aikuisten aineisto on peräisin henkilöiltä, jotka ovat osallistuneet Yleisiin kielitutkintoihin (YKI). YKI-aineistosta näissä hankkeissa käytettiin osakorpusta, jossa on 667 tekstiä ja 49 000 sanetta. Sekä nuorten että aikuisten aineistoissa on mielipidetekstejä ja muodollisia ja epämuodollisia viestejä. Tehtävänannot pidettiin mahdollisimman samanlaisina, vaikka nuorten kokemusmaailma otettiin huomioon. Nuorilta on koottu myös kertomuksia. Aineistojen koostumuksesta tarkempaa tietoa on Jantusen & Pirkolan artikkelissa (2015). Aineiston määrä vaihtelee jonkin verran tutkimuksesta toiseen. Ensimmäisiä opinnäytteitä tehdessä ei koko korpus ollut vielä käytössä ja joissakin tutkimuksissa on keskitytty esimerkiksi yhden tekstilajin kieleen, mutta suurin osa seuraavassa esitellyistä tuloksista on saatu koko Cefling-aineistoa ja YKI-aineistoa tutkimalla ja vertaamalla.

Kaikki tekstit on kirjoitettu ilman apuvälineitä valvotussa tilanteessa ja ennalta määritellyssä ajassa. Ne (myös YKI-aineistot, jotka oli jo kertaalleen arvioitu YKI-taitotasolle) on arvioitu Eurooppalaisen viitekehyksen taitotasolle A1–C2. Koska eri tekstilajeja ja taitotasoa edustamaan haluttiin sama määrä tekstejä, niitä koottiin huomattavasti enemmän kuin tässä on käytetty. Siksi oli mahdollista jättää perusaineistojen ulkopuolelle tekstejä, joiden taitotasosta oltiin epävarmoja. Arviointi on todettu luotettavaksi tilastollisin ja laadullisin analysein (Huhta ym. 2014).

S2-aineistojen pohjalta on valmistunut kolme väitöskirjaa (Seilonen 2013, Kajander 2013, Mustonen 2015). Loppuvaiheissa on lisäksi kaksi artikkeliväitöskirjaa, joiden artikkelit ovat jo pääosin ilmestyneet tai julkaisuissa arvioitavina (ks. mm. Reiman 2011a, 2011b, 2014 ja Mylläri 2020a, 2020b). Suomen kielen maisterintutkielmia hankkeissa on valmistunut 12. Kaikki nämä sekä muutammat julkaisemattomat konferenssiesitelmät ovat tämän artikkelin pohjana. Mukana on myös pari tutkimusta, joissa tarkastellaan vain käyttötaajuutta, koska tarkkuuden määrittely kyseisen ilmiön kohdalla on erityisen ongelmallista. Nämä eivät tietenkään ole

mukana artikkelin pääkysymykseen (taajuuden ja tarkkuuden suhde) vastaamisessa, mutta laajentavat kuitenkin kuvaa eri taitotasolla käytettävien konstruktoiden esiintymisestä. Kyseessä on siis meta-analyysi, johon on koottu käyttötaajuuden ja tarkkuuden kehitystä koskevat tulokset lukuisista tutkimuksista. Niitä ei ole esitetty numeraalisesti, koska muutosten suunnan hahmottaminen valtavasta numeromäärästä tekisi lukemisen erittäin haastavaksi. Kaikki numeeriset arvot, joihin luvun 4 taulukot perustuvat, ovat tietenkin löydettävissä alkututkimuksista.

## 4 Tutkittujen konstruktoiden käyttötaajuuden ja tarkkuuden suhde

### 4.1 Taulukoiden ja alalukujen esittely

Tämän luvun taulukoihin on koottu kaikki tiedossani olevat edellä esiteltyihin aineistoihin perustuvat tutkimukset, joissa on selvitetty jonkin konstruktion käyttötaajuutta ja/tai tarkkuutta. Lisäksi on mukana pari tutkimusta, joissa seurataan sanaston kehitystä. Monia konstruktioita tai muita kielellisiä ilmiöitä on tietenkin tutkimatta, joten mitenkään kattava tämä esittely ei ole. Olen jakanut tutkitut konstruktiot kolmeen alaryhmään: erilaisia lauseita käsittelevät (taulukko 1), lauseensisäisiä konstruktioita käsittelevät (taulukko 2) ja merkityslähtöiset (taulukko 3) tutkimukset. Koska konstruktio jo määritelmänsä nojalla sisältää muodon, merkityksen ja käytön, jaottelu on keinotekoinen ja sen tavoitteena onkin vain helpottaa laajan aineiston hahmottamista. Sanastotutkimukset esitellään erikseen. Mikään ei tietenkään estä etsimästä kehityskulkujen välisiä eroja ja yhteyksiä myös näiden ryhmien jäsenten välillä. Keskityn taitotasojen väliseen vertailuun enkä kommentoi systemaattisesti aikuisten ja nuorten eroja, joita mainitsen lähinnä pohdintaosiossa. En myöskään käsittele tässä erojen mahdollisia syitä tai vaikutuksia. Niitä on jonkin verran käsitelty taustalla olevissa tutkimuksissa ja hiukan tämän luvun pohdinnassakin.

Taulukoissa < merkitsee käytön tai virheiden lisääntymistä, > niiden vähenemistä. Jos muutosta ei ole ollenkaan tai se on suhteellisesti erittäin vähäinen, merkkinä on =. Merkinnyt <= tai >= tarkoittavat pientä muutosta, kaksi nuolta (>>, <<) taas huomattavan suurta muutosta. Joissakin tutkimuksissa on laskettu konstruktoiden tarkkuutta, useimmissa virheitä. Yhdenmukaisuuden vuoksi olen tässä käyttänyt termiä *virhe* ja merkinnyt nuolet sen mukaan. Kun virheet vähenevät, tarkkuus paranee.

### 4.2 Lauseet

Taulukkoon 1 on koottu tietoja niistä tutkimuksista, joiden kohteena ovat erityyppiset lauseet.

TAULUKKO 1. Käytön ja tarkkuuden kehitys erilaisissa lauseissa.

			A1–A2	A2–B1	B1–B2	B2–C1	C1–C2	
Eksistentiaali- lauseet	Aikuiset	Käyttö	<	>=	>=	>	>	
(Kajander 2013)		Virheet	=	>	>	=	>	
	Nuoret	Käyttö	<	=	>=			
		Virheet	>	>	>			
Transitiivi- lauseet	Aikuiset	Käyttö	>	<<	>	<	>	
(Reiman 2011)		Virheet	ei tietoja					
(Reiman 2014)	Nuoret	Käyttö	=	<	<			
		Virheet	>	>	>			
Passiivilauseet	Aikuiset	Käyttö	=	<	<	<<	=	
(Seilonen 2013)		Virheet	>	<	>	>	=	
	Nuoret	Käyttö	<	<				
		Virheet	<	<				
Nollapersoonalauseet	Aikuiset	Käyttö	>	<	<<	>>	>	
(Seilonen 2013)		Virheet	laskettu konstruktio- tyypeittäin, vaihtelee					
	Nuoret	Käyttö	>	<				
		Virheet	laskettu konstruktio- tyypeittäin, vaihtelee					

Kaikkia tutkittuja lausekonstruktioita käytetään A1 tasolta alkaen. Eksistentiaali-lauseiden määrän suunta on laskeva jo A2-tasolta alkaen, kun taas transitiivikonstruktioit lisääntyvät voimakkaasti B1-tasolla ja sen jälkeen niiden osuus alkaa laskea. Passiivilauseiden määrä lisääntyy vähitellen, voimakkaimmin B2- ja C1-tasojen välillä. Nollapersoonalauseiden käytön huippu osuu tasolle B2. Erilaisten lauseiden kirjo siis lisääntyy kielitaidon kehittyessä, jolloin alkuvaiheessa yleisemmät lauseet vähenevät. Eksistentiaali- ja transitiivilauseiden virheiden määrä laskee tasaisesti. Passiivilauseiden virheet lisääntyvät B1-tasolle saakka ja vähenevät sen jälkeen.

### 4.3 Lauseen sisäiset rakenteet

Taulukkoon 2 on koottu tiedot tutkimuksista, joiden kohteena on jokin lauseen osana esiintyvä konstruktio.

TAULUKKO 2. Lauseen sisäisten rakenteiden käytön ja tarkkuuden kehitys

			A1–A2	A2–B1	B1–B2	B2–C1	C1–C2
Negaatio	Aikuiset	Käyttö	<	>	=	>	>
(Halttunen 214)		Virheet	>	>	>>	>	>
	Nuoret	Käyttö	<	>	>		
		Virheet	>	>>	<		
<i>Olla</i> -rakenteet*)	Aikuiset	Käyttö		>		>	
(Kynsijärvi 2007)		Virheet					
Verbiketjut	Aikuiset	Käyttö	<	>	>	=	>
(Paavola 2008)		Virheet	=	>	>	>	>
Rektioilmaukset	Aikuiset	Käyttö	=	=	<	<<	<<
(Nucci 2013)		Virheet	<	>	>	>	>
	Nuoret	Käyttö	<	<	<		
		Virheet	>	>	<		
U-verbit	YKI	Käyttö	=	<	<	<<	<<
(Siitonen & Martin 2012)	LAS2	Käyttö				=	
	Nuoret	Käyttö	<	<	<		
Ilmisubjekti	Nuoret	Käyttö	>	>	>	>	>
(Törö 2017)	(Topling)						
Possessiivi-suffiksi**)	Aikuiset	Käyttö	>	<	<	=	<<
(Turunen 2012)		Virheet	>>	<	>	<	=

Tässä tutkitut ilmiöt ovat luonteeltaan erilaisia. Kieltokonstruktioita ja *olla*-verbiä on miltei mahdotonta välttää. Rektioilmauksia ja verbiketjuja tulee vastaan, kun verbi-valikoima alkaa laajeta. U-verbistä voi vältellä pitkäänkin. Ilmisubjekti taas on tyyli-

kysymys: *minä* on korostainen kirjoitetussa kielessä, kun taas puheessa tai epämuodollisessa viestissä *mä* on usein neutraalimpikin kuin pelkkä verbin taivutusmuoto. Myös possessiivisuffiksia voi pitää tyylikysymyksenä, vaikka joissakin konteksteissa se on välttämätön (esim. *mielelläni*).

Kieltokonstruktioiden käyttö kasvaa aluksi, mutta alkaa sitten vähetä, kun sanavaraston kasvaessa löytyy muitakin keinoja kielteisen asenteen ilmaisuun. Virheiden määrä vähenee jyrkimmin aikuisilla tasojen B1 ja B2 välillä, nuorilla yhtä tasoa aiemmin. Virheet eivät tosin jakaudu tasaisesti eri aikamuotojen kesken: kielteisen imperfektin kaikista virheistä aikuisilla yli puolet on tasolla A2 (Martin 2010).

*Olla*-verbin käyttö vähenee verbivalikoiman kasvaessa. Käyttö on kuitenkin erilaista eri tasoilla. A-tasolla korostuvat yksiverbiset (*on*) tai kiellon sisältävät rakenteet (*ei ole*). B-tasolla kasvaa liittomuotojen osuus, C-tasolla taas lisääntyvät muut moniverbiset *olla*-rakenteet (*on tehtävä, olla nukuksissa* jne.).

Verbiketjuja, joita voidaan pitää merkinä kielitaidon kompleksistumisesta (Mylläri, tulossa), esiintyy jo tasolla A1, mutta ne ovat hyvin prototyyppisiä ja leksikaalista variaatiota on tuskin lainkaan. Verbiketjujen suhteellinen määrä alkaa B1-tasolla vähentyä, mutta niiden pituus ja sisäinen variaatio kasvaa. Virheet vähenevät tasaisesti.

Rektioilmausten käyttö kasvaa eniten C-tasolle tullessa ja vielä C1- ja C2-tasojen välilläkin. Niiden runsaus on siis merkki korkeasta taitotasosta. Nuorilla käyttö kasvaa tasaisesti, mutta virheiden määrä lisääntyy myös B-tasolla.

Ilmisubjektin käyttö vähenee tasaisesti tasolta toiselle. Sen osalta tarkkuuden arviointi on ongelmallista, sillä kieliopillisestihan sen käyttö ei ole virhe. Ilmisubjektien määrässä on kuitenkin vaihtelua semanttisten verbiluokkien välillä. Tutkimus poikkeaa muista tässä käsitellyistä töistä siinä, että sen aineistona on Topling-hankkeen laajempi aineisto, jossa on mukana kirjoittajia alakoululaisista lukiolaisiin.

Possessiivisuffiksien käyttö lisääntyy A2-tasolta alkaen. Jyrkin nousu on tasojen C1 ja C2 välillä. Virheet vähenevät eniten A1 ja A2-tasojen välillä, mutta lisääntyvät käytön lisääntyessä. Vielä C1-tasollakin tarkkuus on hieman alle 80 %, joten possessiivisuffiksien kirjakielen mukainen käyttö on merkki korkeasta kielitaidon tasosta.

Tässä esiteltyjen konstruktioiden lisäksi Ukkola (2009) on selvittänyt samasta aineistosta substantiivilausekkeiden kehitystä, mutta vain kokonaisuutena tasoilla A, B ja C (ei siis A1 vs. A2 jne.). Hän ei myöskään ole esitellyt tarkkuutta, mutta käytön osalta näyttää siltä, että substantiivilausekkeiden määrä ja niissä esiintyvien määritteiden monipuolisuus kasvavat, mutta määritteiden luku lauseketta kohti ei juuri muutu. Kiinnostavinta on, että joitakin määritetyyppejä esiintyy aikuisten C-tasolla jopa kymmenkertainen määrä B-tasoon nähden. Monipuolinen valikoima erilaisia määritteitä näyttäisi siis olevan C-tason indikaattori.



Ajanilmaukset (Varis 2010)	Nuoret	Käyttö Tarkkuus	<	=	<		
			=	<	=		
Veloitteen ilmaiseminen (Kekäläinen 2015)	Aikuiset	Käyttö Tarkkuus	=	=	=	=	=
			=	=	=		

Taulukkoon 3 on koottu tutkimukset, joissa tutkitaan käytön ja tarkkuuden kehitystä jonkin merkityksen ilmaisemisen näkökulmasta. Näistä laajin on Mustonen (2015), jossa selvitetään paikallissijojen käyttöä ja tarkkuutta konkreettisten paikkojen (Spat = spatiaalinen kenttä) ja abstraktimpien tilojen (Circ = sirkumstantiaalinen kenttä) ilmaisussa. Englannin prepositio *at* viittaa tässä paikkafunktioon ja *to* ja *from* väyläfunktioon Mustosen teoreettisen viitekehyksen, konseptuaalisen semantiikan (Jackendoff 1983) käytänteiden mukaisesti. (Mustonen 2015: 99–104.) Yleiskuvana on, että Spat-käytöt vähenevät ja Circ-käytöt lisääntyvät kielitaidon kehittyessä. Erityisesti erosijojen abstrakti käyttö (Circ-FROM) näyttää lisääntyvän melko myöhään. Kummassakin ryhmässä tarkkuus paranee viimeistään A2-tasolta alkaen.

Seilosen (2013) epämääräiseen persoonaan viittaamista tutkivasta työstä geenerisen *sinän* ja *ihmisen* osuus on sijoitettu tähän kategoriaan, koska niissä ei ole kyse uuden lausetyypin oppimisesta. Kumpikaan ei ole kovin tavallinen ja niiden käyttö on enimmillään B2-tasolla.

Nuorten ajanilmaukset ovat enimmillään B2-tasolla, jolla myös tarkkuus on vaikiintunut. Aikuisien teksteistä ajanilmauksia ei ole tutkittu.

Veloitteen ilmaiseminen ja sävyttäminen ei kokonaisuudessaan juuri määrällisesti muutu taitotasolta toiselle, mutta keinovalikoima lisääntyy ja sävyttäjät monipuolistuvat. *Haluta* painottuu alemmilla tasoilla (vrt. Lesonen ym. 2020). Tulkinnaltaan epäselviä tapauksia esiintyy taitotasoilla A1–B1.

## 4.5 Sanasto

Konstruktioihin tai merkityksiin keskittyvien tutkimusten lisäksi Cefling-hankkeen piirissä on tehty kaksi tutkimusta sanaston kehityksestä. Puhakka (2010) tarkastelee *mennä-* ja *tulla-*verbien käyttöä. Aikuisilla se lisääntyy B2-tasolle saakka, nuorilla se alkaa vähentyä jo A2-tasolla. Kummassakin ryhmässä virheet alkavat vähentyä jo A2-tasolta alkaen. Ainoa sanastoa, erityisesti sen rikkautta, yleisemmin selvittänyt tutkimus Cefling-aineistosta on Malin (2012), jonka mukaan sanaston kehitys on hyvin vähäistä tasojen A1 ja A2 välillä. Suuri hyppäys sanaston rikkaudessa on tasojen B2 ja C1 välillä. Sanaston tarkkuuden kehitystä ei ole työssä tarkasteltu. Kaiken kaikkiaan nuorten ja aikuisten S2-sanaston kehityksestä on hyvin vähän tietoa. Kieliympäristön



vaikutusta venäjänkielisten aikuisten sanaston rikkauteen on tutkinut Gronskyte (2020) ja vironkielisten oppijoiden verbisanastoa taitotasolla A2–B2 Taimisto (2014). Alakouluikäisten lasten sanaston kehityksestä on laaja tutkimus (Honko 2013). Sen tulokset ovat merkittäviä, mutta niitä ei voi vertailla tässä esiteltyjen tutkimusten tuloksiin erilaisen aineiston ja tutkimusmenetelmien vuoksi.

#### 4.6 Koontia ja vertailua

Cefling-hankeessa pyrittiin luomaan yleiskuvaa funktionaalisesti määriteltyjen kielitaitotasojen ja joidenkin konstruktoiden tai muiden kielen piirteiden yhteyksistä. Kyse on näennäispitkittäistutkimuksesta ja aineistoja käsitellään yhtenä taitotaso-kohtaisena massana eikä yksittäisten kirjoittajien kehitystä seuralla. Joissakin tutkimuksissa on tosin mainintoja siitä, kuinka monessa eri tekstissä jokin ilmiö esiintyy. Saadut tulokset ovat siis yleistyksiä eikä niiden tilastollisia merkitsevyyttä ole systemaattisesti testattu. Lesonen (2013) tarkasteli kuitenkin siihenastisten opinnäytteiden tuloksia myös tältä kannalta. Hänen mukaansa kieltoa, ajanilmauksia, nollapersoonaa ja *mennä-* ja *tulla-* verbejä käytetään suunnilleen saman verran taitotasosta riippumatta. Nollapersoonaa lukuun ottamatta nämä konstruktiot ovat jotakuinkin välttämättömiä joka tasolla. Impersonaalista passiivista käytetään tilastollisesti merkitsevästi enemmän C-tasolla. Transitivi-ilmausten käytön kasvu on merkitsevää aikuisilla B1-tasolla, mutta ei koululaisilla. Tarkkuus paranee melko tasaisesti kaikkien piirteiden osalta ja laahaa kaikkien piirteiden osalta frekvenssin kasvun jäljessä.

Tarkkuuden kehitystä on selvitetty myös Brunnin ym. (2019) artikkelissa, joka perustuu virheannotoituun Kansainvälisen oppijansuomen korpuksen (ICLFI). Siihen on koottu ulkomaisissa yliopistoissa suomea opiskelevien kirjoituksia eli kyse on suomesta vieraana kielenä. Artikkelissa verrataan erityyppisten virheiden määrää potentiaalisten virheiden määrään. Ilmiöiden käyttötaajuutta ei ole selvitetty, mutta tietenkin potentiaalisten virhepaikkojen määrä kertoo epäsuorasti myös siitä. Tutkimuksessa havaittiin neljä eri tarkkuuden kehityskulkua. Yleisimmin virheet vähenevät taitotasolta toiselle. Sama pitää paikkansa tämänkin katsauksen tutkimuksissa kokonaisuutena. Toiseksi yleisintä on, että virheet lisääntyvät tasolta A2 tasolle B1 ja vähenevät tasolta B1 tasolle B2. Myös tämä on tavallista tässä esiteltyjen tutkimusten pohjalta etenkin aikuisilla. Suomalaista yläkoulua käyvien nuorten osalta samanlainen kehityskulku tapahtuu usein yhtä taitotasoa aiemmin. Virheet saattavat myös vähetä A2-tasolta B1-tasolle ja lisääntyä sitten B2-tasolle. Neljäs mahdollisuus on, että mitään kehityskulkua ei ole havaittavissa. Kaikki nämä vaihtoehdot ovat löydettävissä myös Cefling-tutkimuksista, joten ainakaan aikuisten osalta oppimisympäristö (toinen kieli vs. vieras kieli) ei mitenkään radikaalisti vaikuta mahdollisiin oppimisen reitteihin.

YKI-aineistoon perustuvasta Haapalan (2008) verbiytimiä koskevasta tutkimuksesta käy ilmi, että rakenteellisesti mutkikkaampien verbiydinten määrä kasvaa kohti

ylintä tasoa, virheiden kokonaismäärä vähenee, mutta taivutusvirheiden, vartalo-  
virheiden ja leksikaalisten virheiden keskinäinen suhde ei juuri muutu taitotasojen  
välillä. Tulokset ovat odotuksenmukaisia, mutta eivät suoraan verrattavissa edellä  
esitelyihin erilaisten lähtökohtien vuoksi.

Kun jonkin kielen oppimista aletaan tutkia, huomio kiinnittyy usein ensin virhei-  
siin, joten etenkin vanhempaa virheanalyttistä tutkimusta eri kielistä on valtavasti.  
Lähtökohdiltaan melko lähellä tätä katsausta on Thewissen 2015, jossa tarkastellaan  
ranskankielisten englanninoppijoiden kirjoitettujen tekstien tarkkuutta taitotasolla  
B1–C2. Vaikka taitotasolta löytyy yleisiäkin ilmiöitä, päätulos kuitenkin on, että op-  
pijoiden kirjoittamisen tarkkuus ei edisty lineaarisesti.

## 5 Muna vai kana?

### 5.1 Kielen kehityksen ja tämän katsauksen luonteesta

Kielitaidon kehitys on siis aina yksilöllistä ja epälineaarista. Asioita opitaan ja unoh-  
detaan – sitähän käyttöpohjaisuus perimiltään tarkoittaa. Oppimisen poluista on  
kuitenkin mahdollista löytää myös yhteisiä piirteitä. Tässä tutkitut ilmiöt ovat enim-  
mäkseen sellaisia, että ilman niitä on vaikea kirjoittaa suomen kieltä luontevasti, jos  
ollenkaan. Tekstiä ei ole ilman lauseita, lauseensisäisiä konstruktioita ja sanoja. A1-  
tasollakin kirjoittajat ainakin tavoittelevat monenlaisia lauseita, vaikka yksinkertais-  
ten peruslauseiden osuus onkin suurin: lausetyyppien variaatio kasvaa taitojen ke-  
hittyessä. Myös vaikkapa kieltoja tai paikan ja tilan ilmaisuja tarvitaan joka vaiheessa.  
Niiden käytön ja tarkkuuden muutokset kertovat siten kielitaidon kehityksestä, sillä  
ne osuvat jokaisen oppijan polulle.

Toisen kielen oppimisen peruskysymyksiä on, milloin jokin ilmiö voidaan katsoa  
opituksi (ks. esim. Hyltenstam & Abrahamsson 2003). Oikeiden muotojen prosentua-  
aalinen osuus (ks. luku 2.2) tai sanavaraston laajuus ovat usein käytettyjä mittareita,  
mutta yhtä tärkeää on havainnoida, miten ja millaisia kielen piirteitä oppija käyttää  
tai ymmärtää. Sääntöpohjaisen kielenoppimisenäkemyksen mukaan sääntö katso-  
taan opituksi vasta, kun sitä sovelletaan useissa konteksteissa. Silloin oppijan tuotta-  
maa ilmausta ei voida pitää ulkoa osattuna kokonaisuutena, jollaisten käytön ei siis  
ajatella olevan oikeastaan oppimista (esim. Pienemann 1998). Käyttöpohjainen vii-  
tekehys muuttaa tätä oppimisen määrittelyä, sillä siinä konstruktion ensimmäisen-  
kin ymmärrettävän esiintymän katsotaan osoittavan oppimista ja muodostavan läh-  
tökohdan, josta konstruktio voi lähteä toistumaan, varioimaan ja abstrahoitumaan.

Vaikka etenkin varhaisimpien oppijankielen ilmiöiden luokittelu on usein haas-  
tavaa, tutkittavan konstruktion käyttötaajuus tuhatta sanaa kohti voidaan laskea  
melko yksiselitteisesti, kunhan konstruktion rajat on huolellisesti määritetty. Tarkkuus

on ongelmallisempi ilmiö. Tämän katsauksen pohjana olevissa tutkimuksissa tarkkuus määrittyy eri tavoin sen mukaan, mihin tutkimus kohdistuu. Tutkimustehtävän rajaus määrittelee, painottuuko ilmauksen hyväksyttävyyden muodon vai käytön tai merkityksen tarkkuuteen, vaikka konstruktion määritelmä sisältää kaikki nämä puolet. ICLFI-korpuksen virheluokitus (Brunni ym. 2019) painottuu vahvasti muodon virheisiin, vaikka muutama merkitykseen liittyvä kategoriakin on mukana. Tarkkuutta käsittelevät tutkimustulokset eivät siis ole täsmällisesti verrattavissa tutkimuksesta toiseen, vaikka jonkinlainen intuitio kohdekielen mukaisuudesta ja hankkeiden piirissä käydyt laajat keskustelut ja yhteiset luokittelusessiot toki lisäävät yhdenmukaisuutta.

Tämä katsaus ei perustu tarkkoihin laskelmiin vaan summittaisiin pienten (>, <) ja suurten (>>, <<) muutosten arviointiin osittain luokitteluongelmien vuoksi, osittain siksi, että kaikin kohdin tarkkoja lukuja ei ole käytettävissä. Aineistokaan ei ole kovin suuri, etenkin nuorten B2-tasolla. Verrattuna muihin kielenkehitystä havainnoiviin tutkimuksiin, joissa usein tyydytään vertaamaan ryhmien kielitaitoa esimerkiksi käytyjen kurssien tai opiskeluajan pohjalta, aineiston vahvuus lienee sen taitotasoarvioinnissa. Arvioinnissa oli mukana suuri joukko ihmisiä, joista osa oli arvioinnin ammattilaisia ja muutkin enimmäkseen hyvin kokeneita arvioijia. Yksityiskohdista ja valinnoista keskusteltiin paljon. Arvioinnin onnistumista tutkittiin jälkepäin sekä määrällisesti että laadullisesti (Huhta ym. 2014) ja se todettiin niin luotettavaksi kuin funktionaalinen taitotasoarviointi voi ylipäänsä olla.

Käyttöraajuuden ja tarkkuuden suhteeseen laajemman näkökulman tuo Schmidin Entrenchment-and-Conventionalization-Model (2020), jonka mukaan kielisysteemi sekä yksilön että yhteisön tasolla syntyy konstruktioiden juurtumisen ja vakiintumisen myötä kohti konventionaalista kielenkäyttöä. Vaikka tässä esiteltyjä hankkeita suunniteltaessa Schmidin mallia ei vielä ollut, taustaoletuksissa on paljon yhteistä ja olisi jopa mahdollista ajatella tässä esitettyjä tuloksia jonkinlaisena Schmidin mallin implementaationa.

## 5.2 Päätulokset

Kielen ilmiön käyttöraajuus ja tarkkuus voivat periaatteessa olla sidoksissa toisiinsa kolmella tavalla: ne voivat kasvaa yhtä matkaa funktionaaliselta taitotasolta toiselle, toinen voi taantua tai pysyä samana toisen kasvaessa tai niiden välillä ei ole mitään suhdetta. Hyvin karkeasti yleistäen voi tähänastisen tutkimuksen pohjalta sanoa, että välttämättömimmät konstruktiot ovat käytössä alusta alkaen riippumatta niiden teoreettisesta mutkikkoudesta. A1-tasolla ne voivat esiintyä hyvin suppeasti, kuten vain yhdessä kontekstissa tai vain harvoin leksikaalisiin yksiköihin rajoituen: esimerkiksi verbiketjut sisältävät vain pari tavallisinta verbiparia (esimerkiksi *osaa puhua, haluaa lähteä*). Tavallisimpien konstruktioiden tarkkuus paranee jo A2-tasolla, etenkin nuorilla. B1-tasolle on leimallista erilaisten konstruktioiden käytön

lisääntyminen, usein jonkinlainen ilmaisun vuolaus. Osataan ilmaista monenlaista, mutta usein aika epätarkasti. Tosin B1-tasolle voi päästä myös melko kapealla kielellisten ilmausten valikoimalla, jos osaa niiden avulla suorittaa annetun tehtävän, ja tällöin tarkkuus voi olla myös melko hyvä. Kummassakin tapauksessa B1 taso todentuu itsenäisen kielenkäytön tasoksi: tehtävästä selviydytään ja ymmärrettävyys on riittävä. B2-tasoa leimaa tarkkuuden yleinen paraneminen: B1-tasolla käyttöön tulleet konstruktiot tarkentuvat. C-tasolla taas kasvaa harvinaisempien konstruktioiden käyttö, vaikka ne voivat olla vielä epätarkkoja. Ilmeinen C-tason indikaattori voisi olla myös sanaston rikastuminen, mutta tästä ei suomen kielessä ole vielä juuri tutkimusta (ks. kuitenkin Malin 2012).

Vaikka käyttötaajuus ja tarkkuus kehittyvät usein yhtä matkaa, näiden tutkimusten valossa on selvää, että taitotasoportailta käyttötaajuuden kasvu edeltää tarkkuuden paranemista. Jos siis käyttö on kana ja tarkkuus muna, kana tulee ensin.

### 5.3 Hyvät ja pahat virheet

Virheitä pelätään ja vältellään. Kielenoppimisessa ne kuitenkin ovat välttämättömiä: kielitaito edistyy kieltä käyttämällä. Vasta kehkeytymässä olevat konstruktiot rajautuvat vähitellen kohdekielen mukaisiksi, kun niitä käytetään ja saadaan palautetta. Virheet ovat siis oppimisen edellytys. Rohkeasti kokeileva oppija edistyy parhaiten (ks. esim. Lesonen 2020: 132).

Virheitä pidetään usein heikon kielitaidon ominaisuutena. Niin tietenkin onkin: kielitaidon edistyessä tarkkuus paranee. On kuitenkin myös virheitä, jotka kertovat korkeatasoisesta kielitaidosta. Niitä voi kutsua positiivisiksi virhetypeiksi (Thewissen 2015: 212). Ne voivat olla tyyli- tai rekisterivirheitä (Brunni ym. 2019) tai liittyä konstruktioidiin, jotka tyyppillisesti opitaan melko myöhään, koska niitä on mahdollista kiertää käyttämällä jo tuttuja ilmaisutapoja (esim. U-verbit, monet rektiota-paukset, passiivirakenteet, possessiivisuffiksit). Tällöin virhe kertoo siitä, että oppija ei mene yli siitä, missä aita on matalin, vaan pyrkii rikastamaan ilmaisuun.

Tällaiset havainnot voivat olla hyödyllisiä myös opetuksen ja kielitaidon arvioinnin kannalta. Aiemmin kielenopetus lähti usein tarkkuudesta: uusi asia opetettiin perin pohjin ensin sääntöjen ja esimerkkien avulla ja vasta sitten sitä alettiin käyttää omista tuotoksissa. Tarkkuuden siis ajateltiin olevan käytön edellytys. Käyttöpohjainen näkemys kielenoppimiseen taas väittää, että kielitaito syntyy käytöstä ja tarkkuus kehittyy käytön myötä. Käyttöä ja tarkkuutta havainnoimalla voi tehdä päätelmiä oppijan kehitysvaiheesta, esim. Brunni ym. (2019) ehdottaa, että syntaktisen ja morfosyntaktisen tarkkuuden lisääntymisen voisi nostaa yhdeksi tasoa B2 määrittäväksi piirteeksi. Eurooppalaisen viitekehyksen funktionaalinen lähtökohta ja sovellettavuus kaikkiin kieliin ei tietenkään mahdollista yksittäisten kielten piirteiden nostamista yleisiin taitotasokuvauksiin, mutta niiden rinnalle voidaan tutkimuksen ja kokemuksen lisääntyessä vähitellen hahmotella kielikohtaisia tietoja

siitä, mikä eri tasoille on tyypillistä. Kielen rakenteiden ohella myös sanastoa olisi S2-alalla tutkittava tästä näkökulmasta nykyistä enemmän.

Kirsti Siitosen keskeinen tutkimusaihe, U-verbit ja niihin kietoutuvat muut kielen keinot, joiden avulla ilmaisun suoruutta sekä tekijän näkymistä ja intentiota voidaan säädellä, on nostanut esiin erityisesti edistyneiden kielenoppijoiden kirjoittamisen haasteet. Hän pani alulle myös LAS2-korpuksen, jonka avulla tämän ryhmän kielen piirteistä on saatu ja saadaan lisää tietoa. Tämän artikkelin näkökulmasta U-verbitutkimus nosti esiin tärkeän seikan: kielitaidon kehitys voi näkyä sekä uusien konstruktoiden hallintana ja niiden käytön kasvuna että niihin liittyvien virheiden lisääntymisenä. Hallintaahan ei voi osoittaa kieltä käyttämättä ja uuden ilmiön kokeilu ja haltuunotto johtaa väistämättä aluksi myös virheisiin. Juuri mitään kielen ilmiötä ei ole edes mahdollista kuvata ja selittää niin täydellisesti, että oppija voisi edes teoriassa käyttää sitä aina oikein ensi esiintymästä lähtien. Oikeat käytöt syntyvät kokeilemalla ja siihen kielenoppijoita tulee rohkaista.

## Kirjallisuus

- Brunni, S., L.-M. Lehto, J.H. Jantunen, & V. Airaksinen 2015. How to annotate morphologically rich learner language. Principles, problems and solutions. Teoksessa A-K. Helland Gujord, S. Nacey & S. Ragnhildstveit (toim.) *Learner corpus research: LCR2013 conference proceedings*. Bergen language and linguistics studies 6. Bergen: University of Bergen, 133–152. <https://doi.org/10.15845/bells.v6i0.812>.
- Brunni, S., J.H. Jantunen & V. Skantsi 2019. Korpusavusteinen virheanalyysi tarkkuuden kehityksestä EVK:n taitotasolla A2-B2. *Puhe ja kieli*, 39 (3), 275–304. <https://journal.fi/pk/article/view/76601>.
- Bybee, J.L. & C. Beckner 2015. Usage-Based Theory. Teoksessa *The Oxford Handbook of Linguistic Analysis* (2 ed.) B. Heine & H. Narrog (toim.). <https://doi.org/10.1093/oxfordhb/9780199677078.013.0032>.
- Cefling: The Linguistic Basic of the Common European Framework levels 2007–2009. <https://www.jyu.fi/hytk/fi/laitokset/kivi/tutkimus/hankkeet/paattyneet-tutkimushankkeet/cefling/suom>.
- Common European Framework of Reference for Languages (CEFR). <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions>.
- Ellis, N.C. 2016. Frequency in language learning and language change. Teoksessa *Experience Counts: Frequency Effects in Language*. H. Behrens & S. Pfänder (toim.). Boston: De Gruyter. <https://doi.org/10.1515/9783110346916-010>.
- Eskildsen, S.W. 2009. Constructing another language—Usage-Based Linguistics in Second Language Acquisition. *Applied Linguistics*, 30 (3), 335–357, <https://doi.org/10.1093/applin/amn037>.
- EVK = Eurooppalainen viitekehys: *Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys* 2003.
- Franceschina, F., R. Alanen, A. Huhta & M. Martin 2006. A progress report on the Cefling project. Invited paper presented at SLATE (Second Language Acquisition and Testing in Europe) Workshop, University of Amsterdam.

- Glahn, E., G. Håkansson, B. Hammarberg, A. Holmen, A. Hvenekilde & K. 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, A. 1995. *Constructions. A construction grammar approach to argument structure. Cognitive theory of language and culture*. Chicago, IL: The University of Chicago Press.
- Gronskyte, A. 2020. *Kieliympäristön vaikutus S2-oppijoiden sanaston laajuuteen*. Pro gradu -tutkielma. Turun yliopisto. <http://urn.fi/URN:NBN:fi-fe20201217101261>.
- Haapala, T. 2008. *Finiittiverbeistä verbiketjuihin: verbiytymien kompleksistuminen S2-oppijoiden kielessä*. Pro gradu -tutkielma. Tampereen yliopisto. <https://urn.fi/urn:nbn:fi:uta-1-17666>.
- Halttunen, S. 2014. *“Sori mä en voin tullan” : negaatioilmaukset suomi toisena kielenä -oppijoiden teksteissä*. Pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:jyu-201410132977>.
- Honko, M. 2013. *Alakouluikäisten leksikaalinen tieto ja taito: toisen sukupolven suomi ja S1-verrokki*. Acta Universitatis Tampereensis 1865. Tampere: Tampereen yliopisto. <http://urn.fi/URN:ISBN:978-951-44-9251-8>.
- Housen, A., F. Kuiken & I. Vedder 2012. *Dimensions of L2 Performance and Proficiency. Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins.
- Huhta, A., R. Alanen, M. Tarnanen, M. Martin & T. Hirvelä 2014. Assessing learners' writing skills in a SLA study: Validating the rating process across tasks, scales and languages. *Language Testing*, 31, 307–328. <https://doi.org/10.1177/0265532214526176>.
- Hyltenstam, K. & N. Abrahamsson 2003. Age of onset and ultimate attainment in near-native speakers of Swedish as a second language. Teoksessa K. Hyltenstam & K. Fraurud (toim.) *Multilingualism in Global and Local Perspectives*. Stockholm: Stockholm University & City of Stockholm, 319–340. <https://www.researchgate.net/publication/303813117>.
- Ivaska, I. 2015. *Edistyneen oppijansuomen konstruktiopiirteitä korpusvetoisesti: avainrakennanalyysi*. Turun yliopiston julkaisu, C 409. Turku: Turun yliopisto.
- Jackendoff, R. 1983. *Semantics and cognition*. Cambridge, MA: MIT.
- Jantunen, J. H. 2011. Kansainvälinen oppijan-suomen korpus (ICLFI): typologia, taustamuuttujat ja annotointi. Lähivördlusi. *Lähivertailuja*, 21, 86–105. Tallinna: Eesti Rakenduslingvistika Ühing.
- Jantunen, J. H., S. Brunni, L.-M. Lehto & V. Airaksinen 2014. Oppijankieliaineistojen annotointi. Esimerkkinä ICLFI:n annotoinnin prosessit, ongelmat ja ratkaisut. Teoksessa M. Mutta, P. Lintunen, I. Ivaska & P. Peltonen (toim.) *AfinLA-e: Soveltavan kielitieteen tutkimuksia*, 7. Jyväskylä: Suomen soveltavan kielentutkimuksen yhdistys, 60–80.
- Jantunen, J. H. & S. Pirkola 2015. Oppijansuomen sähköiset tutkimusaineistot. Nykytilanne. *Virittäjä*, 119 (1). <https://journal.fi/virittaja/article/view/46508>.
- Kaivapalu, A. & Martin, M. 2018. DEMfad-malli–apua tutkimustulosten vertailuun? Esitelmä Kielitieteen päivillä Helsingissä 7.–9.5.2018.
- Kajander, M. 2013. *Suomen eksistentiaalilause toisen kielen oppimisen polulla*. Jyväskylä Studies in Humanities 220. Jyväskylä: Jyväskylän yliopisto. <http://urn.fi/URN:ISBN:978-951-39-5530-4>.
- Kekäläinen, M. 2015. *Velvoitteen ilmaiseminen ja sävyttäminen aikuisten suomenoppijoiden teksteissä*. Pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:jyu-201506102274>.
- Kitsnik, M. 2018. *Iga asi omal ajal: B1- ja B2-tasemekeelekonstruktsioonid keeleoskuse taseme näitajana*. Tallinna: Tallinnan yliopisto.

- Kynsijärvi, T. 2007. *Se johtuu siitä, että minulla oli muistinmenetys: olla-verbirakenteiden kehkeytyminen oppijankielessä*. Pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:ju-201302241264>.
- Lesonen, S. 2013. *Suomen kielen lingvististen piirteiden sijoittuminen eurooppalaisen viitekehyksen taitotasolle*. Pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:ju-201302241264>.
- Lesonen, S. 2020: *Valuing Variability: Dynamic Usage-based Principles in the L2 Development of Four Finnish Language Learners*. Groningen Dissertations in Linguistics. <http://urn.fi/URN:ISBN:978-94-034-2675-4>.
- Lesonen, S., R. Steinkrauss, M. Suni & M. Verspoor 2020. Lexically specific vs. productive constructions in L2 Finnish learners. *Language & Cognition*, 12 (3), 526–563.
- Malin, E. 2012. *Suomi toisena kielenä -oppijoiden sanaston kehittyminen taitotasolta toiselle siirryttäessä*. Pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:ju-201211223060>.
- Martin, M., S. Mustonen, N. Reiman M. Seilonen 2010: On Becoming an Independent User. Teoksessa I. Bartning, M. Martin & I. Vedder 2010 *Communicative proficiency and linguistic development, intersections between SLA and language testing research*. EUROSLA Monograph Series 1. European Second Language Association, 57–80. <http://www.eurosla.org/monographs/EM01/EM01home.php>.
- Martin, M. 2010. Connecting functional proficiency levels with linguistic development: experiences from the Cefling Project in Finland. Kutsuttu esitelmä Lundin yliopistossa 16.11.2010.
- Mustonen, S. 2015. *Käytössä kehittyvä kieli: paikat ja tilat suomi toisena kielenä -oppijoiden teksteissä*. Jyväskylä Studies in Humanities 255. Jyväskylä: Jyväskylän yliopisto. <http://urn.fi/URN:ISBN:978-951-39-6213-5>.
- Mylläri, T. 2020a: Words, clauses, sentences, and T-units in learner language: Precise and objective units of measure? *Journal of the European Second Language Association*, 4 (1), 13–23. <https://doi.org/10.22599/jesla.63>.
- Mylläri, T. 2020b: Measuring syntactic complexity in learner Finnish. *Apples – Journal of Applied Language Studies*, 14 (2), 67–92. <https://doi.org/10.47862/apples.99134>.
- Nucci, A. 2013. *Me voimme nauttia luontoa: rektioilmaukset suomi toisena kielenä -oppijoiden teksteissä*. Pro gradu -tutkielma. Jyväskylän yliopisto.
- Paavola, V. 2008. *Haluaisitko menna muunkansa kalastaman? Verbiketjujen kehkeytyminen suomi toisena kielenä -oppijoiden kielessä*. Pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:ju-2011062911091>.
- Pienemann, M. 1998. *Language Processing and Second Language Development. Processability Theory*. Studies in Bilingualism 15. Amsterdam: John Benjamins.
- Puhakka, M. 2010. *”Sit se meni ja tuli hetken päästä takas.” Verbit mennä ja tulla suomi toisena kielenä -oppijoiden teksteissä*. Pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:ju-201007132299>.
- Reiman, N. 2011a. Transitiivikonstruktio ikkunana syntaksin kehitykseen: passiivi ja infiniittiset rakenteet taidon indikaattoreina S2-oppijoiden teksteissä. Teoksessa E. Lehtinen, S. Aaltonen, M. Koskela, E. Nevasaari & M. Skog-Södersved (toim.) *AFinLA-e: soveltavan kielitieteen tutkimuksia*, 3, 142–157. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinla/article/view/4465>.
- Reiman, N. 2011b. Two faces of complexity: structural measures and diversity of constructions. *Nordand: Nordisk tidsskrift for andrespråksforskning*, 6 (2), 9–33.
- Reiman, N. 2014. Yläkoulun S2-oppilaiden transitiivi-ilmausten käyttö Eurooppalaisen viitekehyksen taitotasolla. *Lähivördlusi. Lähivertailuja* 21, 183–220. <http://dx.doi.org/10.5128/LV24.07>.



- Ruuska, K. 2020. *At the nexus of language identity and ideology. Becoming and being a highly proficient second language speaker of Finnish*. Jyväskylä Studies in Humanities 309. Jyväskylä: Jyväskylän yliopisto. <http://urn.fi/URN:ISBN:978-951-39-8366-6>.
- Schmid, H.-J. 2020. *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford: Oxford University Press.
- Seilonen, M. 2013. *Epäsuora henkilöön viittaaminen oppijansuomessa*. Jyväskylä Studies in Humanities 197. Jyväskylä: Jyväskylän yliopisto. <http://urn.fi/URN:ISBN:978-951-39-5025-5>.
- Siitonen, K. & M. Martin 2012. Suomen kielen U-verbit taitotason indikaattorina. *Lähivördlusi. Lähivertailuja*, 22, 369–406. <http://dx.doi.org/10.5128/LV22.13>.
- Siivelt, K. & S. Mustonen 2013. Lähdekielen vaikutus ja kielitaitotasot: paikallissijojen kehitys oppijansuomessa. *Lähivördlusi. Lähivertailuja*, 23, 341–370. <http://dx.doi.org/10.5128/LV23.14>.
- Skehan, P. 1998. *A Cognitive Approach to Language Learning*. Oxford, Oxford University Press.
- Skehan, P. 2009. Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*, 30 (4), 510–532. <https://doi.org/10.1093/applin/amp047>.
- Spoelman, M. & M. Verspoor 2010. Dynamic Patterns in Development of Accuracy and Complexity: A Longitudinal Case Study in the Acquisition of Finnish. *Applied Linguistics*, 31 (4), 532–553. <https://doi.org/10.1093/applin/amq001>.
- Taimisto, H. 2014. *Taitotasolta toiselle: korpuspohjainen tutkielma vironkielisten suomenoppijoiden verbisanaston kehittymisestä*. Pro gradu -tutkielma. Oulun yliopisto.
- Thewissen, J. 2015. *Accuracy across proficiency levels: A learner corpus approach*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Tilma, C. 2014. *The dynamics of foreign versus second language development in Finnish writing*. Jyväskylä Studies in Humanities 233. Jyväskylä: Jyväskylän yliopisto.
- Tomasello, M. 2003. *Constructing a Language. A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Topling = Toisen kielen oppimisen polut 2010–2013. <https://www.jyu.fi/hytk/fi/laitokset/kivi/tutkimus/hankkeet/paattyneet-tutkimushankkeet/topling>.
- Turunen, U. 2012. *Suomi toisena kielenä -oppijoiden possessiivisuffiksien käyttö*. Pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:ju-201206131859>.
- Törö, M. 2017. *Ilmisubjektin esiintyvyys kirjoitetussa oppijansuomessa*. Pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:ju-201704202023>.
- Ukkola, A. 2009. Taas yksi nollatutkimus: substantiivilausekkeiden määritteet S2-oppijoiden teksteissä. Pro gradu -tutkielma.
- Varis, K. 2010. *Ajanilmaukset Cefling-hankkeen koululaisaineistossa*. Pro gradu -tutkielma. Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:ju-201009302818>.