VENÄLÄISTEN SUOMEN OPISKELIJOIDEN OPPIJANSUOMESTA

Suomen kielen
kandidaatintutkielma
Oulun yliopisto
10.1.2020

Katariina Halonen

SISÄLLYS

1. JOHDANTO

Tutkin Venäjällä asuvien, venäjää äidinkielenään puhuvien suomen opiskelijoiden suomea. Tutkin suomen kieltä heidän kirjoittamistaan opiskelutehtävistä, jotka on kerätty sähköiseen korpukseen. Selvitän eri sanaluokkien määrät teksteissä taitotasoilla B1 ja B2 ja paljonko niitä on suhteessa koko kyseisen taitotason sanemäärään. Lisäksi selvitän, muuttuuko niiden osuus kielitaidon kehittyessä eli siirryttäessä tasolta B1 tasolle B2, ja jos muuttuu, niin mihin suuntaan. Kyseessä on kvantitatiivinen perustutkimus, jonka tuloksia voidaan hyödyntää myöhemmässä tutkimuksessa vertailuaineistona. Aihe kiinnostaa minua, sillä olen aikaisemmin opiskellut venäjää ja asunut Venäjällä.

Aineistoni on Kansainvälisestä oppijansuomen korpuksesta, International Corpus of Learner Finnish (ICLFI). Se on kirjoitetun kielen korpus, johon on kerätty tekstejä yli 20 ulkomaisesta yliopistosta, joissa opetetaan suomea (Jantunen & Pirkola 2015: 92). Korpuksessa on kaikkiaan 959 840 sanetta (Jantunen, Brunni & Oulun yliopisto: 2013). Korpuksen tekstit on taitotasoarvioitu eurooppalaisen viitekehyksen mukaan.

Tutkimusongelmani on kuinka sanaluokkien määrät jakautuvat eri taitotasoilla oppijansuomessa ja toisaalta, miten niiden määrät muuttuvat eri taitotasoilla vai muuttuvatko ne.

Tutkimuskysymykseni ovat seuraavat:
1) Miten eri sanaluokkien määrät jakautuvat eri taitotasoilla?
2) Miten jakautuminen muuttuu kielitaidon kehittyessä?

Sanaluokkien osalta oppijansuomen tutkimusta ei ole juurikaan tehty aikaisemmin. Sen sijaan oppijankieltä on tutkittu ja ICLFI-korpuksen teksteistä on tehty useita tutkimuksia. Marianne Spoelman (2013) on tutkinut väitöskirjassaan partitiivin käyttöä oppijansuomessa sellaisten opiskelijoiden osalta, joiden äidinkieli on viro, saksa ja hollanti. Hanna Varrio (2014) on tutkinut pro gradu -tutkielmassaan *sanoa*-verbiä fraseologisena yksikkönä oppijansuomessa ja natiivisuomessa. Olli-Juhani Piri (2017) ja Helena Kuuluvainen (2015) ovat tutkineet pro gradu -tutkielmissaan oppijansuomen virheitä. Piri tutki objektivirheitä ja Kuuluvainen fraseologisia virheitä. Myös Maria Huttu-Hiltunen (2017) ja Annika Liiti (2018) ovat tarkastelleet pro gradu -tutkielmissaan oppijansuomen virheitä. Huttu-Hiltunen tutki virheitä virolaisten suomenoppijoiden joka-, mikä- ja kuka-

relatiivikonstruktioissa. Liiti käsitteli ruotsinkielisten suomenoppijoiden morfosyntaktisten verbimuotojen virhekäyttöä. Suomenoppijoiden kielitaidon kehitystä taitotasoilla siirryttäessä ovat tutkineet Sanna Mustonen (2015) ja Mikko Kajander (2013) Jyväskylän yliopistossa. Mustonen tutki paikkoja ja tiloja suomenoppijoiden teksteissä eri taitotasoilla ja Kajander tutki suomen kielen eksistentiaalilausetta ja taitotasoja.

Tutkielmani etenee niin, että luvussa kaksi esittelen tutkimukseni teoreettista taustaa ja metodia sekä käsittelen suomen kielen sanaluokat. Sen jälkeen luvussa kolme käsittelen tutkimusaineistoani, kansainvälistä oppijansuomen korpusta, ja yleisesti venäläisten tekstejä korpuksessa. Luvussa neljä analysoin aineistoa, kerron kuinka analyysi etenee, käsittelen sanaluokkien osumat korpuksesta taitotasoittain ja teen yhteenvedon tuloksista. Luku viisi on tutkielman päätäntö.

2. TUTKIMUKSEN TEOREETTINEN TAUSTA JA METODI

Tässä luvussa esittelen teoriaa tutkimukseni taustalla, tutkimusmetodia ja lopuksi käyn myös läpi suomen kielen sanaluokat, joita tarkastelen aineistonani olevasta oppijansuomesta. Ensin käsittelen korpuksia ja korpuslingvistiikkaa, sen jälkeen oppijankieltä ja kielen oppimisen yhteistä eurooppalaista viitekehystä. Sen jälkeen kerron oppijankielen korpustutkimuksesta ja kontrastiivisesta välikielen analyysistä, tutkimusmetodistani. Lopuksi esittelen suomen kielen sanaluokat, joiden pohjalta tutkin aineistoa.

2.1. Korpustutkimus, oppijankieli ja kontrastiivinen välikielen analyysi

Laajasti ajateltuna **korpus** tarkoittaa suurta kieliaineistoa, jonka avulla tarkastellaan sitä kielivarianttia, jota kyseinen korpus edustaa (Ivaska 2015: 49). Kieltä tai kieliä tutkivalle korpus tarkoittaa yleensä puhuttua tai kirjoitettua, elävässä elämässä käytettyjen kielen esimerkkien kokoelmaa. Se on valittu edustavasti kielestä kokonaisuutena tai joistakin kielellisistä genreistä. Useimmiten modernit korpukset, jotka ovat elektronisia korpuksia, sisältävät vähintään miljoona sanaa, ja niissä on kokonaisia tekstejä tai pitkien tekstien laajoja osia. Tekstit ovat keskenään erilaisia, mutta yhteistä niille on se, että ne koostuvat yhden kielen tai yhden kielen varieteetin tai rekisterin teksteistä. Korpuksiksi kutsuttavien aineistojen kirjo on laaja, vaikka niille asetetaankin tiettyjä kriteereitä, kuten edustavuus, autenttisuus, koneluettavuus, riittävä kokopysyvyys, julkisuus ja dokumentoituus. (Lounela & Heikkinen 2012: 121.) Korpuksiin lisätään useimmiten annotaatioita eli kulloisenkin korpuksen kannalta mielekästä metatietoa. Tällaista tietoa voi olla esimerkiksi tekstin ulkopuoliset taustamuuttujat, kuten tuottamispaikka, julkaisukanava tai kirjoittaja, tekstin sisäinen kappale- tai virkerakenne tai esimerkiksi metalingvistinen annotaatio, kuten morfologisten muotojen, syntaktisten funktioiden tai sanaluokkien merkitseminen. (Ivaska 2015: 49.)

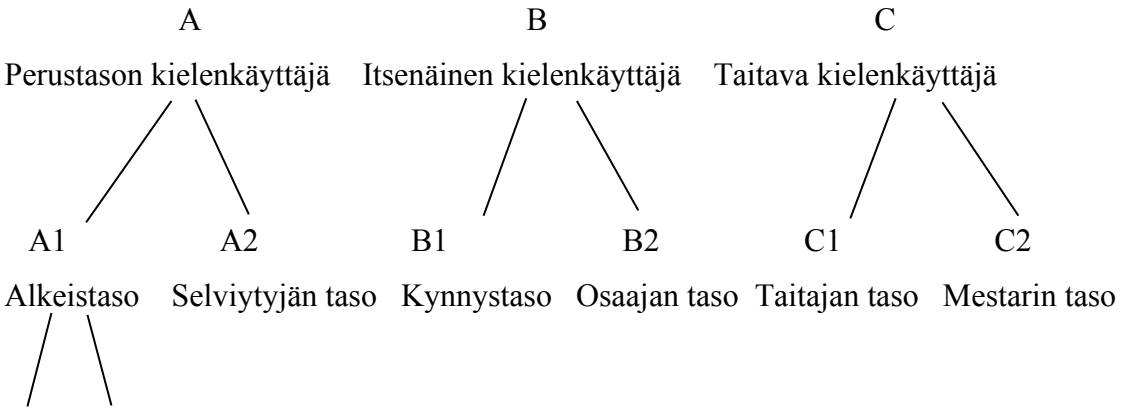
Korpuslingvistiikka on tutkimusta, jossa hyödynnetään kielellisiä korpuksia. Sitä pidetään tutkimusmetodien joukkona eikä niinkään kielenkuvauksen teoriana. Siinä tarkastellaan aitoa kielenkäyttöä, suuria aineistoja, suhteellisia taajuuksia ja erilaisia yhdistelmiä,

kuten jonkin piirteen ja tekstityypin välisiä yhdistelmiä tai sanojen ryhmien välisiä yhdistelmiä. (Lounela & Heikkinen 2012: 121–122.) Korpuslingvistiikan perimmäisenä metodologisena periaatteena voidaan pitää sitä, että tarkoituksena on kuvata kielessä vallitsevia yleisiä lainalaisuuksia suuren tekstimassan avulla (Jantunen 2004: 32).

Korpuslingvistiikassa tutkimuksen kohteena voi olla myös **oppijankieli**. Se on yksi nimitys toisen tai vieraan kielen oppijan kielelle. Muilla oppijan osaamilla kielillä voi olla vaikutusta oppijankieleen, mutta ennen kaikkea oppijankieli syntyy oppijan lähtökielen, opittavan kielen ja kulloisenkin oppijankielen vaiheen yhteisvaikutuksessa. Oppijankielellä voidaan nähdä olevan erilaisia vaiheita, ja se myös muuttuu koko ajan: opitaan uusia sääntöjä ja vanhoja muokataan. Oppijankielen nähdään olevan itsenäinen järjestelmä lähde- ja kohdekielen välissä. Sen voidaan sanoa olevan systemaattinen ja luova vaihe kielenoppimisessa, sen piirteet ovat ainutkertaisia. Oppijan kielimuodossa esiintyy poikkeamia verrattaessa kohdekieleen, mutta ne nähdään virheiden sijaan oppijankielen ominaisuuksina. (Nissilä 2011: 39–40.) Oppijankieli on jatkuvasti alttiina muutoksille, ja se muuttuu nopeasti, mikä on seurausta mm. siitä, ettei oppija kuule ympärillään omaa kielimuotoaan vastaavaa kieltä. Yleisesti ottaen oppijankieli kehittyy yksinkertaisesta kompleksiseen, sen systeemi laajenee kaikilla kielellisillä tasoilla. Oppija omaksuu monimutkaisempia sääntöjä, sääntöjä yleistetään ja toisaalta joitain sanamuotoja ja fraaseja opitaan sellaisenaan. Jotkin muutokset oppijankielessä voivat tapahtua kuitenkin erityisen hitaasti, jolloin kielen kehitys tuntuu pysähtyneen, vaikka niin ei todellisuudessa olisikaan. (Latomaa 1993: 20–21.) Kun opiskeltava kieli on suomi, puhutaan **oppijansuomesta**. Latomaa (1996: 98) pitää oppijansuomea neutraalimpana terminä kuin esimerkiksi välikieli tai vierassuomi, vaikka siihenkin sisältyy kysymyksiä, kuten mistä oppijansuomi alkaa ja milloin henkilön suomi ei ole enää oppijansuomea. Nissilä (2011: 37) käyttää termiä oppijankieli, koska pitää sitä omana kielimuotonaan, ei vain välivaiheena.

Vierasta kieltä opiskeltaessa kielitaidon tasoa voidaan mitata eri tavoin. Euroopan tasolla yhtenäisen arvioinnin avuksi on kehitetty yhteinen **eurooppalainen viitekehys** (EVK). Siinä on määritelty kielitaidon kuvaamisen objektiiviset kriteerit. Siinä kuvataan oppijoiden kielitaidon tasoja, joilla oppijan ajatellaan siirtyvän aina ylöspäin kielitaidon kehittyessä. Viitekehyksessä määritellään ne tasot, joiden avulla kielitaidon edistymistä voidaan mitata oppimisen aikana. Kielitaito on jaettu kolmeen laajaan taitotasoon: A, B ja C, jotka

jakaantuvat lisäksi ainakin kahteen tasoon, esimerkiksi A1 ja A2. Kuviossa 1 tasot on esitetty haaroittuvalla mallilla, johon on tarpeen mukaan helppo lisätä tasoja, kuten A1.1. ja A1.2. (EVK 2003: 19, 47, 57.)



KUVIO 1. Kielitaidon taitotasot esitettynä haaroittuvalla mallilla. (EVK 2003: 47.)

Tasoa A nimitetään perustason kielenkäyttäjän tasoksi, jolta löytyy alkeistaso ja selviytyjän taso. Taso B on itsenäisen kielenkäyttäjän taso, joka jakaantuu kynnystasoon ja osaajan tasoon. Tasolla C puhutaan taitavasta kielenkäyttäjästä, ja taso jakaantuu taitajan tasoon ja mestarin tasoon. Tasoille on annettu yleisiä kuvauksia. Peruskielitaitoon eli A-tasoon kuuluu esimerkiksi tuttujen arkipäivän ilmausten ymmärtäminen ja käyttäminen, perustietojen kertominen itsestään, yksinkertainen keskustelu, jos puhekumppani puhuu selvästi ja hitaasti ja on valmis auttamaan. B-taso on itsenäisen kielenkäyttäjän taso, jolla kielenkäyttäjä ymmärtää pääasiassa yleiskieliset viestit niin työssä, koulussa kuin vapaa-ajalla. Hän pystyy tuottamaan yksinkertaista ja johdonmukaista tekstiä, kuvaamaan kokemuksia, toiveita ja tavoitteita. Viestiminen on niin sujuvaa ja spontaania, että vuorovaikutuksessa syntyperäisen kanssa kummankaan ei tarvitse ponnistella. Suomen kansalaisuuden saamiseksi suomen taidon tulee olla vähintään tasolla B1, jota vaadittu yleisen kielitutkinnon taitotaso kolme vastaa (Kansalaisuuslaki § 17). C-tasolla puhutaan taitavasta kielenkäyttäjästä, joka ymmärtää yleensä vaivatta kaikenlaista puhuttua ja kirjoitettua kieltä, pystyy itse puhumaan sujuvasti ja spontaanisti, käyttää kieltä joustavasti ja tehokkaasti niin sosiaalisissa tilanteissa kuin opintoihin ja ammattiinkin liittyvissä tilanteissa. (EVK 2003: 47–48.)

Taitotasojen lisäksi kielitaito on jaettu kuullun ymmärtämiseen, luetun ymmärtämiseen, puheen tuottamiseen ja kirjoittamiseen. Jokaisella tasolla ja kielitaidon osa-alueella on omat kuvauksensa. Kirjoittamisen osalta kehitys kulkee niin, että A-tasolla pystytään kirjoittamaan lyhyitä ja yksinkertaisia tekstejä, kuten postikortti lomaterveisin, lyhyt arkinen muistiinpano tai täyttämään omat henkilötiedot lomakkeeseen. B-tasolla kirjoitetaan yhtenäistä, yksinkertaista tekstiä tutuista aiheista. Myös kirjoitelman tai raportin kirjoittaminen onnistuu ja siinä välitetään tietoa tai esitetään perusteluja. C-tasolla osataan valita tekstiin sopiva tyyli ja kirjoitetaan jäsennettyjä, yksityiskohtaisia ja monipolvisia tekstejä. Kirjoituksessa osataan ottaa lukija huomioon ja auttaa tätä löytämään ja muistamaan tekstistä tärkeimmät asiat. Tietyn alan sanaston käyttö on sujuvaa. (EVK 2003: 34–35, 50–51.)

Seuraavaksi palaan takaisin korpuksiin ja käsittelen oppijankielen korpustutkimusta ja tutkimusmetodiani kontrastiivista välikielen analyysia.

Oppijankielen korpustutkimus on alkujaan ollut ennen kaikkea englannin kielen tutkimusta, mutta aineistoja on syntynyt runsaasti myös muista kielistä, ja oppijankieliaineistot ovat maailmanlaajuisesti kasvava korpusaineistojen muoto. Muita kohdekieliä ovat nykyään muun muassa espanja, ranska, italia, saksa, arabia, unkari, viro ja suomi. Suurin osa aineistoista on kirjoitetun kielen aineistoja. (Jantunen & Pirkola 2015: 88.) Oppijankielen korpukset lasketaan usein erikoiskorpuksiksi (Ivaska 2015: 50). Erikoiskorpukset eivät kuvaa niin sanottua tavallista kieltä, sillä ne sisältävät kyseiselle kielelle epätavallisia piirteitä tai kieli ei ole luonnollisessa ympäristössä käytettyä. Erikoiskorpusten voi sanoa keskittyvän erityisesti opetusprosessiin ja olevan erityisen hyödyllisiä virheiden analysoinnissa. (Tognini-Bonelli 2001: 8–9) Grangerin (2002: 13) mukaan oppijankielen korpustutkimuksen käytännön päämäärä liittyykin vieraiden kielten opetukseen ja oppijoiden kielitaidon parantamiseen. Oppijankielen korpuksiin kuuluvat oleellisena osana erilaiset metatiedot. Ne ovat tietoja oppijoista, heidän oppimisensa kontekstista ja korpukseen liitetyistä tuotoksista. Oppijansuomen kohdalla tällaisia ovat yleisesti esimerkiksi oppijan äidinkieli, suomen taitotaso, kauanko on opiskellut suomea, käyttääkö suomea kotona ja mitä muita vieraita kieliä oppija hallitsee. (Jantunen & Pirkola 2015: 94–95.)

Korpustutkimukseen liittyy metodina kiinteästi **kontrastiivinen välikielen analyysi** (Contrastive Interlanguage Analysis, CIA). Siihen sisältyy sekä natiivikielen ja ei-natiivikielen vertailu että ei-natiivikielen ja ei-natiivikielen vertailu. Jälkimmäinen tarkoittaa oppijankielen eri varianttien vertailua. Natiivikielen ja ei-natiivikielen, esimerkiksi natiivisuomen ja oppijansuomen, vertailun avulla voidaan löytää oppijoiden kielestä niin sanottujen virheiden lisäksi myös sanojen, lausekkeiden ja rakenteiden yli- ja aliedustumista. Grangerin mukaan tätä vertailua voidaan tehdä, vaikka oppijankieli nähtäisiinkin omana kielisysteeminään. Vertaaminen natiivikieleen ja natiivikielen normeihin mahdollistaa siitä poikkeavien esiintymien yleisyyden arvioimisen. Grangerin mukaan on oleellista verrata oppijankieltä natiivikieleen, sillä kaikki vieraiden kielten opetus tähtää oppijoiden osaamisen saamiseen lähemmäs natiivikieltä. (Granger 2002: 8–9, 12–13.) Oppijankielen eri varianttien vertailu voi olla esimerkiksi oppijankielen vertailua eri äidinkielisten oppijoiden välillä. Voidaan myös vertailla sellaisten oppijoiden oppijankieltä, joilla on sama äidinkieli. Näin esimerkiksi tässä tutkimuksessa, jossa vertaillaan venäjänkielisten oppijankieltä eri taitotasoilla. Granger (2002: 13) toteaa kontrastiivisen välikielen analyysin sisältävän myös vertailun oppijankielen eri varianttien välillä. Se mahdollistaa paremman ymmärryksen oppijankielestä. Vertailut eri äidinkielisten oppijoiden välillä antavat tietoa siitä, mitkä piirteet esiintyvät useilla oppijaryhmillä ja mitkä taas esiintyvät vain tietyn äidinkielen kanssa.

Seuraavaksi esittelen lyhyesti suomen kielen sanaluokat, jotka ovat tutkimukseni kohteena.

2.2. Suomen kielen sanaluokat

Suomen kielen sanaluokkajako perustuu sille, että samaan sanaluokkaan kuuluvat sanat käyttäytyvät olennaisilta osilta samalla tavalla. Taivutus on keskeinen sanaluokkakriteeri. Sen mukaan sanat jakautuvat kolmeen pääryhmään, jotka ovat nominit ja verbit sekä kolmantena ryhmänä adpositiot, adverbit ja partikkelit yhdessä. Nomineihin kuuluvat substantiivit, adjektiivit, pronominit ja numeraalit, joilla kaikilla on sija- ja lukutaivutus. Verbeillä taas on persoona-, modus- eli tapaluokka- ja tempus- eli aikamuototaivutus. (ISK

2004 § 438.) Esittelen seuraavaksi järjestyksessä ensin nominit, sitten verbit ja lopuksi adpositiot, adverbit ja partikkelit.

Nomineja ovat substantiivit, adjektiivit, pronominit ja numeraalit. Substantiiveihin kuuluvat yleisnimet ja erisnimet. Ne nimeävät esineitä, olioita, paikkoja, ominaisuuksia, suhteita, toimintoja ja asiaintiloja. Substantiiveja ovat esimerkiksi *tietokone, kirjasto, kauneus, jäätelö, Vilma, onni* ja *syöminen*. Substantiivit on avoin luokka, niitä tulee jatkuvasti lisää. Adjektiivit ilmaisevat ominaisuuksia, luonnehtivat asioita, olioita, asiaintiloja tai tilanteita. Adjektiiveja ovat esimerkiksi *pyöreä, ihana, suuri* ja *viimeinen*. Adjektiivit on myös avoin luokka, siihen tulee uusia sanoja lähinnä johdoksina ja lainoina. Pronominit sen sijaan eivät kuvaile tai luokittele vaan niiden avulla käsitellään tarkoitteita luokittamatta. Pronomineihin kuuluvat esimerkiksi *hän, tuo, mikä, joka* ja *itse kukin*. Numeraaleista käytetään myös nimitystä lukusanat eli ne ilmaisevat lukumäärää. Numeraaleja on periaatteessa ääretön määrä, niitä voidaan muodostaa lisää peruslukujen ja kymmenkantaisten peruslukujen pohjalta lisää. Näiden lisäksi numeraaleihin kuuluvat myös murtoluku- ja likimääräilmaukset. Numeraaleihin kuuluvat esimerkiksi *kaksi, yksitoista, puoli* ja *kymmenkunta*. (ISK 2004 § 551, 603, 713, 770.)

Verbit kuvaavat tyypillisesti tapahtumia ja tekoja, kuten *välähtää* tai *heittää*. Ne voivat myös kuvata epädynaamisia tilanteita, kuten *rakastaa* tai *koostua*. Verbit ovat avoin luokka, johon tulee jatkuvasti uusia sanoja erityisesti johdoksina. (ISK 2004 § 445.)

Adpositiot, adverbit ja partikkelit taas muodostavat pääasiassa taipumattomien sanojen ryhmän. Adpositioihin kuuluvat prepositiot ja postpositiot. Ne ilmaisevat jonkin asian tai olion suhteen johonkin toiseen. Näin esimerkiksi *alla* tai *keskellä* ilmauksissa *tuolin alla* ja *keskellä päivää*. Ensimmäisessä kyseessä on postpositio, sillä adpositio on täydennyksensä jäljessä, toisessa taas prepositio, koska adpositio on täydennyksensä edellä. Adpositioihinkin voi tulla lisää sanoja eli se on avoin luokka. Adverbit ilmaisevat muun muassa aikaa, paikkaa, olotilaa, tapaa ja määrää. Tällaisia ovat esimerkiksi *illalla*, *koulussa, yhtäkkiä* ja *vähän.* Partikkelit ovat aina taipumattomia ja ne eivät saa määritteitä. Partikkeleilla on useita eri alaryhmiä, muun muassa interjektiot, kuten *oho, kääk*; konjunktiot, kuten *että* ja *vaikka* sekä intensiteettisanat, kuten *aivan* ja *kovin.* (ISK 2004 § 687, 646, 792.)

3. TUTKIMUSAINEISTO

3.1. Kansainvälinen oppijansuomen korpus

Aineistoni on kerätty Kansainvälisestä oppijansuomen korpuksesta, International Corpus of Learner Finnish (ICLFI). Se on kirjoitetun kielen korpus, johon on kerätty suomen opiskelijoiden kirjoittamia tekstejä yli 20 ulkomaisesta yliopistosta. Tekstien kirjoittajat ovat opiskelleet suomen kieltä yliopistossa pää- tai sivuaineena tai yksittäisinä kursseina. (Jantunen & Pirkola 2015: 92). Kaikkiaan korpuksessa on 4 850 tekstiä ja 959 840 sanetta (Jantunen, Brunni & Oulun yliopisto: 2013). Tekstit on taitotasoarvioitu eurooppalaisen viitekehyksen mukaan. Jokaisen tekstin on arvioinut vähintään kaksi eri arvioijaa. Jos teksti on saanut näiltä kaksi eri arviota, on pyydetty vielä kolmas. Taitotaso kuvaa siis jokaisen tekstin tasoa, ei opiskelijan tasoa. (Jantunen, Brunni, Lehto & Airaksinen 2014: 67.) Korpuksen teksteistä suurin osa on taitotasoilta B1 ja B2. Tason B1 tekstejä on 43 % ja tason B2 tekstejä 36 % korpuksen kaikista teksteistä. Tason C1 tekstejä on 12 % ja tason A2 tekstejä 7 %. A1-tasolta tekstejä on alle prosentti ja tasolta C2 2 %. (Jantunen & Pirkola 2015: 93; Jantunen ym. 2014: 66.)

Korpuksen teksteihin on lisätty monenlaista metadataa, kuten tekstien tuottajiin, teksteihin ja keräystilanteeseen liittyvää tietoa. Niihin on myös lisätty lingvististä tietoa, joka selittää tekstiä ja sen elementtejä. Tällaista tietoa on esimerkiksi sanan sanaluokka tekstikontekstissaan. Tätä erilaisten tietojen lisäämistä kutsutaan annotoinniksi. Oppijankielen korpuksissa erityisen oleellinen annotointitaso on sanan lemmatisointi. Siinä sanan taivutettuun muotoon lisätään tieto sanan perusmuodosta eli lemmasta, esimerkiksi **kodussa* 'kodissa' > KOTI. Näin hakemalla sanan perusmuotoa saadaan haettua kaikki sanan taivutusmuodot. ICLFI-korpukseen on tehty myös virheannotaatiota. (Jantunen ym. 2014: 61–63, 71.) Granger (2002: 12–13) mainitsee tietokoneavusteiden virheanalyysin olevan kontrastiivisen välikielen analyysin ohella toinen metodologinen lähestymistapa oppijankielen korpustutkimuksessa.

ICLFI-korpuksen tekstit on lisäksi annotoitu morfologisesti, jolloin sanat ovat saaneet morfologisen koodin. Se on tehty puoliautomaattisesti, mikä tarkoittaa sitä, että automaattisen koodauksen jälkeen morfosyntaktinen koodaus on vielä tarkistettu manuaalisesti

(Jantunen ym. 2014: 68–69). Sanoilla on siis morfologisen roolin mukainen koodi. Morfologisiin rooleihin liittyy myös lisämäärityksiä kuten substantiiveilla, onko kyseessä yksikkö vai monikko ja missä sijassa sana on. (ICLFI-manuaali.) Näin esimerkissä (1) sana *talo*:

(1) *Meillä on **talo** kaupungin ulkopuolella.* (ICLFI: VE0056[1])

Korpuksesta löytyy sanalle *talo* seuraavat tiedot tässä kontekstissa: sen perusmuoto on *talo* ja koodi eli morfologinen analyysi on @NH N SG NOM. Koodissa NH tarkoittaa pääsanaa, N substantiivia, SG yksikköä ja NOM nominatiivia. (ICLFI-manuaali.)

ICLFI-korpuksen teksteihin on merkitty kattavasti metatietoja. Olen konnut ne taulukkoon 1.

[1] Koodi viittaa Kansainväliseen oppijansuomen korpukseen, International Corpus of Learner Finnish, lyhennettynä ICLFI-korpus. Alun kirjainyhdistelmä viittaa oppijan äidinkieleen, tässä tapauksessa VE, joka viittaa venäjään. Sen jälkeen tuleva numero yksilöi oppijan. Mahdollisesti numeron jäljessä oleva kirjain kertoo, monesko kyseisen kirjoittajan teksti tämä on.

TAULUKKO 1. ICLFI-korpuksen tekstien metatiedot.

Keräyspaikka
Keräysvuosi
Oppijan syntymävuosi
Sukupuoli
Syntymäpaikka
Asuinpaikka
Äidinkieli
Vanhempien äidinkielet
Puhutaanko oppijan kotona suomea tai ovatko läheiset opettaneet tälle suomea
Opettajan äidinkieli
Onko oppija ollut Suomessa
Oppijan taitotaso tuntimäärän ja eurooppalaisen viitekehyksen mukaan, ulkopuolisten arvioijien arvioimana
Tekstityyppi
Tehtävänanto
Onko kyseessä ollut koetilanne
Onko kirjoitusaika ollut rajattu
Kirjoituspaikka
Mahdollisesti käytetyt apuvälineet
Tiedostonimi

Metatietojen perusteella pystyn etsimään ICLFI-korpuksen tekstien joukosta juuri niiden oppijoiden tekstit, jotka asuvat Venäjällä, ja joiden äidinkieli on venäjä. Ilman niitä tällainen tutkimus ei olisi mahdollinen. Korpuksessa on mukana myös Virossa asuvia venäjänkielisiä oppijoita, jotka on kuitenkin rajattu tutkimukseni ulkopuolelle. Näin vironkielinen ympäristö ei ole päässyt vaikuttamaan tutkimaani oppijansuomeen ja toisaalta jatkossa on mahdollista verrata Venäjällä ja Virossa opiskeltua oppijansuomea niiden opiskelijoiden osalta, joiden äidinkieli on venäjä.

3.2. Venäläiset kansainvälisessä oppijansuomen korpuksessa

Venäjänkielisten Venäjällä asuvien tekstejä on korpuksessa kaikkiaan 556 kappaletta (ICLFI-manuaali). Taulukossa 2 on kuvattu montako tekstiä ja sanetta korpuksessa on kultakin taitotasolta.

TAULUKKO 2. Venäjänkielisten Venäjällä asuvien tekstit ja saneet taitotasoittain. (ICLFI-manuaali.)

Taitotaso	**Tekstien määrä**	**Saneiden määrä**
A1	1	17
A2	41	3 693
B1	252	42 840
B2	223	71 559
C1	35	16 216
C2	4	3 260
Yhteensä	**556**	**137 585**

Kuten taulukosta näkyy, tekstit keskittyvät taitotasoille B1 ja B2. Tutkimukseni keskittyy näiden taitotasojen tarkasteluun, sillä ei ole luotettavaa vertailla keskenään sellaisia tasoja, joista toisesta on vain yksi tai muutamia tekstejä ja toisesta useampi sata. Myös saneita on muilla tasoilla huomattavasti vähemmän.

Näiden ICLFI-manuaalin lukujen perusteella saa yleiskuvan kaikista Venäjällä asuvien venäjänkielisten teksteistä. Tutkimus on tehty korpuksesta Korp-työkalulla saaduista määristä, minkä takia ne eroavat hieman näistä luvuista.

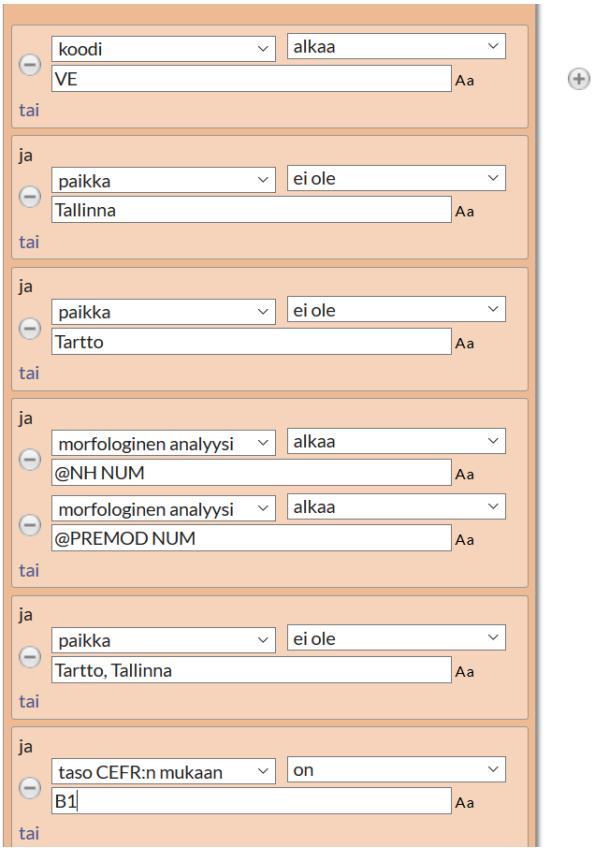
4. ANALYYSI

4.1. Analyysin kulku

Tässä luvussa käsittelen aineiston analyysia. Ensin esittelen, kuinka analyysi etenee eli miten teen haut korpukseen. Sen jälkeen käyn läpi aineistosta löytyvät sanaluokat kaikkien venäläisten osalta ja taitotasoittain.

Korpus on Kielipankissa ja sitä tutkitaan Korp-työkalulla. Tehdessäni hakuja korpukseen Korp-työkalulla minun tulee aina ensin rajata hakuni siten, että mukaan tulevat vain ne tekstit, joiden kirjoittajan äidinkieli on venäjä ja lisäksi rajata haun ulkopuolelle mukana olevat Viron asuinpaikat. Muuten Virossa asuvien, venäjää äidinkielenään puhuvien suomen opiskelijoiden tekstit tulisivat myös mukaan. Lisään ensin hakuun erikseen komennot ”paikka” ”ei ole” ”Tallinna” ja ”paikka” ”ei ole” ”Tartto”. Osumia tarkastellessani huomasin, että mukaan tulee ottaa myös komento ”paikka” ”ei ole” ”Tartto, Tallinna”, eli molemmat samassa, sillä tällaisiakin osumia löytyi.

Äidinkielen ja paikan rajauksen jälkeen lisään hakuun tietyn sanaluokan koodin tai koodit, kuten @NH NUM ja @PREMOD NUM kun etsin numeraaleja. Näillä koodeilla saan haettua sekä pääsanoina, koodissa NH, että määritteinä, koodissa PREMOD, olevat numeraalit. Eri taitotasojen saneita hakiessani lisään hakuun komennon ”taitotaso CEFR:n mukaan” ”on” ja haluamani tason, kuten ”B1”. Alla kuvassa 1 näkyy kuva Korpiin tekemästäni venäläisten tekstien numeraalien hausta taitotasolle B1.



KUVA 1. Venäläisten tekstien numeraalien haku taitotasolta B1.

Kuten kuvasta 1 näkyy, haut koostuvat useasta eri komennosta. Ensinnäkin koodin VE-alulla otetaan mukaan vain venäjänkielisten tekstit ja jokainen Viron paikkavaihtoehto rajataan erikseen haun ulkopuolelle. Lisäksi tiettyjä sanaluokkia haetaan omilla morfologisen analyysin koodeilla ja eri taitotasoja tarkasteltaessa haluttu taso lisätään mukaan omana komentonaan. Eri komentojen välillä on ”ja”, paitsi morfologisen analyysin koodien välillä ”tai”, koska ne eivät voi olla voimassa samaan aikaan. Yksi sane ei voi olla samaan aikaan sekä pääsana että määrite.

Seuraavaksi käsittelen jokaisen sanaluokan erikseen. Kerron, millä morfologisen analyysin koodilla ja mahdollisilla muilla komennoilla sanaluokan sanat löytyvät aineistosta. Ensin haen kunkin sanaluokan saneiden määrän aineistoni venäjänkielisten teksteissä yhteensä ja sen jälkeen kunkin sanaluokan saneiden määrän erikseen sekä taitotasolla B1 että taitotasolla B2. Lisäksi annan esimerkkejä kunkin sanaluokan saneista. Suhteutan taitotasolta saamani tulokset kyseisen tason koko sanemäärään. Tämän avulla pystyn havainnollistamaan, eroavatko sanaluokkien määrät taitotasojen välillä.

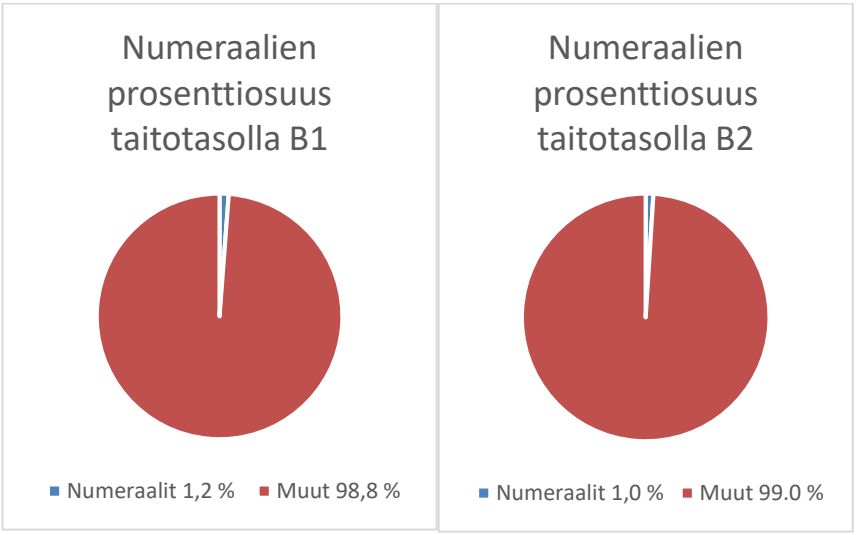
4.2. Sanaluokat taitotasoittain

Seuraavaksi esittelen jokaisen sanaluokan koodin tai koodit ja muut haussa tarvittavat komennot. Joissain tapauksissa tietyn sanaluokan saneita ei löydy suoraan tietyllä koodilla. Näissä tapauksissa kerron, mitä muita toimia tarvitsee tehdä, jotta saan selville kyseisen sanaluokan osumien määrän. Annan esimerkkejä jokaisesta sanaluokasta ja lisäksi kuvaan ympyrädiagrammilla sanaluokan saneiden määrän prosentteina koko taitotason saneista. Tämän esitän kummaltakin tutkimaltani taitotasolta, sekä B1 että B2.

Numeraalit haen koodin alun VE ja Viron paikkojen poissulkemisen lisäksi morfologisen analyysin aluilla @NH NUM ja @PREMOD NUM. Niitä löytyy venäjänkielisten Venäjällä asuvien teksteistä kaikkiaan 1 125. Kun lisään hakuun halutun taitotason, ”komento taso CEFR:n mukaan” ”on” ”B1” ja toisessa haussa ”B2”, saan tulokseksi 370 numeraalia taitotasolle B1 ja 614 numeraalia taitotasolle B2. Mukana on muun muassa peruslukuja, vuosilukuja ja järjestyslukuja, kuten seuraavissa esimerkeissä (2–4):

(2) *Noin **1,5** tuntia me menemme mukavalla bussilla, mutta Mozhgassa meidän bussi oli rikki ja siitä vaihtaan.* (ICLFI: VE0058)
(3) *On olemasa tarina, jonka kirjoitti E. Hoffmann vuonna **1816**.* (ICLFI: VE0002)
(4) ***Ensimmäinen** on flunssasta, **toinen** on rustokoruista ja **kolmas** on tämänhetkistä elämätilanteesta.* (ICLFI: VE0097)

Kuviossa 2 esitän numeraalien määrät kummallakin taitotasolla prosentteina koko kyseisen taitotason saneista.



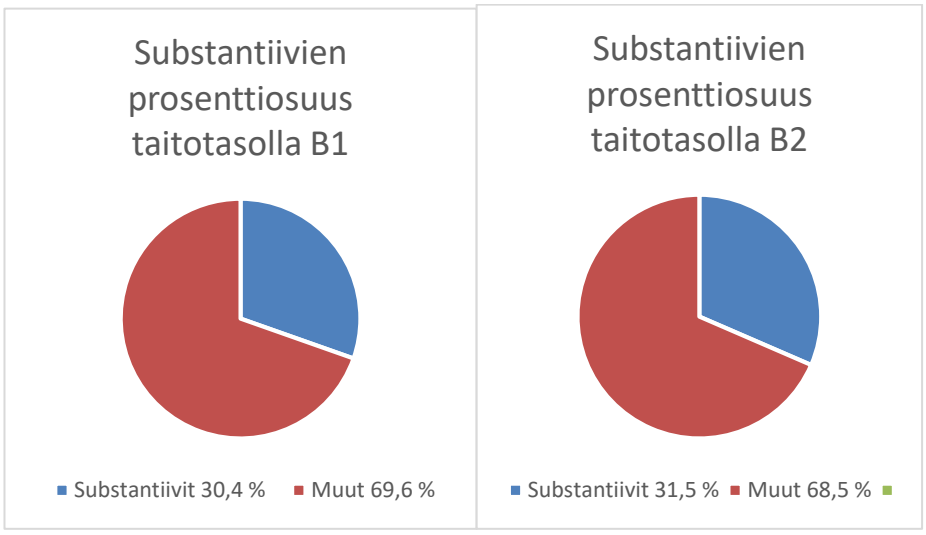
KUVIO 2. Numeraalien prosentuaaliset osuudet kaikista taitotason saneista taitotasoilla B1 ja B2.

Kuten kuviosta 2 näkyy, numeraalien määrä suhteessa taitotason kokonaissanemäärään on 1,2 % taitotasolla B1 ja 1,0 % taitotasolla B2.

Substantiivit haen korpuksesta laittamalla koodin alkamaan VE, rajaamalla Viron paikat haun ulkopuolelle ja lisäksi laitan morfologisen analyysin sisältämään joko koodin @NH N tai @PREMOD N. Tulos on 34 982. Tämän jälkeen vähennän saadusta saneiden määrästä numeraalien määrän, joka on 1 125. Ne tulevat myös mukaan haun tuloksiin, koska numeraalien koodit ovat @NH NUM ja @PREMOD NUM. Lopputuloksena venäläisten teksteissä on kaikkiaan substantiiveja 33 857. Vastaavalla tavalla laskettuna substantiiveja on 9 103 taitotasolla B1 ja 19 470 taitotasolla B2. Esimerkeissä (5–6) substantiivi on pääsanana ja esimerkeissä (7–8) määritteenä:

(5) *Mutta **maine** Jumalalle me suoriutuimme…* (ICLFI: VE0058)
(6) ***Yliopistosta** minä Leenan kanssa lähdimme linja-autoasemaan.* (ICLFI: VE0059)
(7) *Arkangelin kotiseutumuseossa te voitte nähdä Pomorin **elämäntapan** tavarat, pukut ja eläimet.* (ICLFI: VE0051)
(8) *Opettajat tulisivat hiiriksi, jotta he eivät rajoita **opiskelijoiden** vapausta.* (ICLFI: VE0068)

Kuviossa 3 on substantiivien osuudet kummallakin taitotasolla prosentteina koko taitotason sanemäärästä:



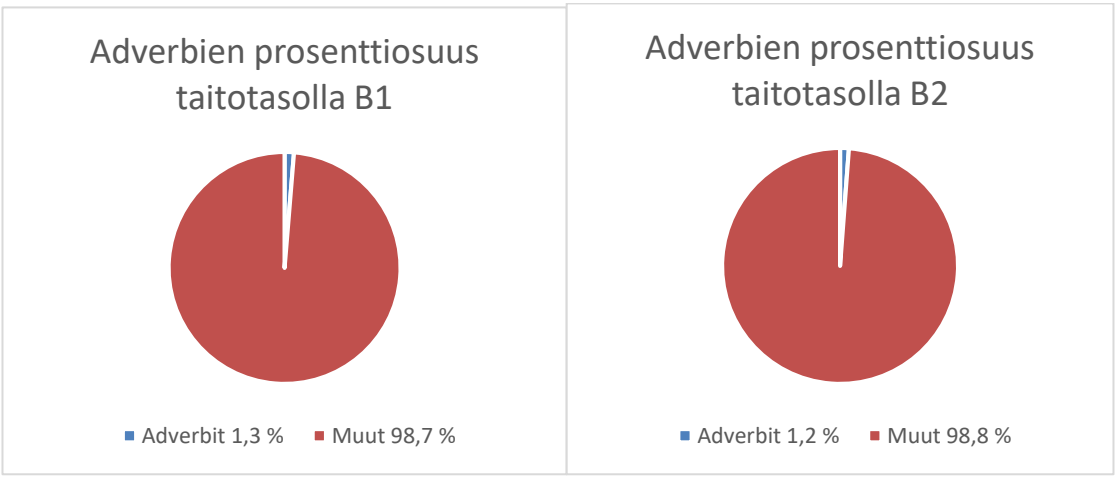
KUVIO 3. Substantiivien prosentuaaliset osuudet kaikista taitotason saneista taitotasoilla B1 ja B2.

Kuten kuviosta 3 näkyy substantiivien osuus kaikista saneista taitotasolla B1 on 30,4 % ja taitotasolla B2 se on 31,5 %. Kummallakin tasolla substantiivien määrä on huomattava, ja tasojen välillä substantiivien prosenttiosuudessa on 1,1 % prosenttiyksikön ero.

Adverbit löytyvät morfologisen analyysin koodilla @PREMOD ADV ja komennolla ”koodi” ”alkaa”. Jos haen komennolla ”koodi” ”on” hausta jää pois mm. tapaukset, joissa adverbi on erisnimessä, kuten elokuvan nimessä *Vähän kunnioitusta*. Tässä tapauksessa adverbin koodi on @PREMOD ADV Prop. Adverbejä on venäläisillä kaikkiaan 1 319. Taitotasolla B1 387 ja tasolla B2 760. Esimerkeistä (9–10) näkyy adverbien käyttöä:

(9) *Se on **aivan** hauskasti.* (ICLFI: VE0124)
(10) *Elämä on **hyvin** vaarallinen rohkeillekin.* (ICLFI: VE0165)

Kuviossa 4 on esitettynä adverbien osuudet prosentteina koko taitotason saneista.



KUVIO 4. Adverbien prosentuaaliset osuudet kaikista taitotason saneista taitotasoilla B1 ja B2.

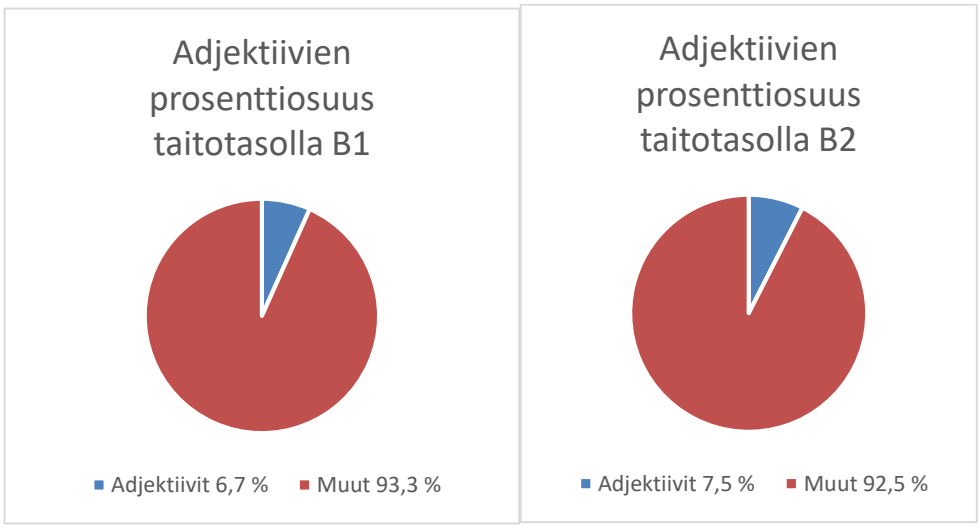
Kuten kuviosta 4 nähdään tasolla B1 adverbien prosenttiosuus kaikista saneista on 1,3. Tasolla B2 se on hieman pienempi 1,2. Tasojen välillä ei siis ole kuin yhden prosenttiyksikön kymmenyksen ero.

Adjektiivien haussa morfologisen analyysin koodit alkavat @NH A ja @PREMOD A. Mukana voi olla myös tietoa, onko adjektiivi yksikössä vai monikossa ja missä sijassa se on. Tämän takia adjektiiveja ei voida hakea komennolla ”koodi” ”on”, koska silloin mukaan tulevat vain sellaiset taipumattomat adjektiivit kuten *koko, viime, eri.* Haen siis komennolla ”koodi” ”alkaa”, jolloin haun tulokseen tulee mukaan myös adverbejä, jotka vähennetään yhteismäärästä. Adjektiivien ja adverbien yhteismäärä kaikilla venäjänkielisillä on 9 247, tasolla B1 2 399 ja tasolla B2 5 399. Vähennysten jälkeen adjektiivien kokonaismääräksi tulee 7 928, taitotasolla B1 määrä on 2 012 ja tasolla B2 määrä on 4 639.

Esimerkissä (11) adjektiivi on pääsanana ja esimerkissä (12) määritteenä:

(11) *Elokuvassa on paljon tyhmiä tapauksia, mutta joskus ne ovat* ***hauskoja****.* (ICLFI: VE0001)
(12) *Nyt ei ole hämmästettävää, että monet ihmiset osaavat monia* ***vieraita*** *kieliä.* (ICLFI: VE0002)

Kuviossa 5 on esitettynä adjektiivien osuudet prosentteina koko taitotason saneista.



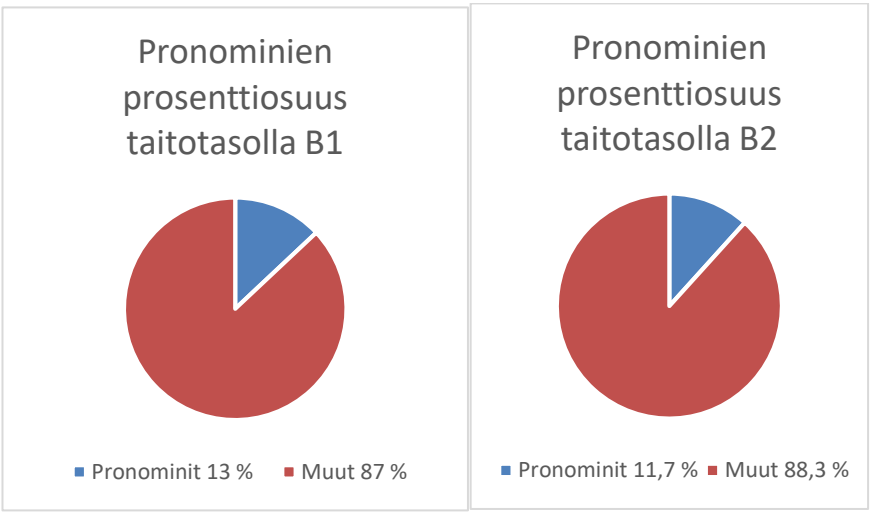
KUVIO 5. Adjektiivien prosentuaaliset osuudet kaikista taitotason saneista taitotasoilla B1 ja B2.

Kuviosta 5 nähdään, että adjektiivien osuus taitotasolla B1 on 6,7 % ja 7,5 % taitotasolla B2. Eroa on 0,8 prosenttiyksikköä, ja osuudessa on hieman kasvua siirryttäessä tasolle B2.

Pronominit löytyvät aineistosta äidinkielen ja paikan rajausten lisäksi morfologisen analyysin koodeilla @NH PRON ja @PREMOD PRON. Niitä löytyy kaikkiaan 12 932, taitotasolla B1 3 899 ja B2 7 216. Mukana on paljon persoonapronomineja, relatiivipronomineja, demonstratiivipronomineja ja kvanttoripronomineja. Esimerkeissä (13–14) pronomini on pääsanana ja esimerkissä (15) määritteenä:

(13) ***Tästä*** *lähtien seikkailut alkoivat.* (ICLFI: VE0001)
(14) ***Minä*** *katsoin televisiota joka päivä, jotta tietäisin uutisia.* (ICLFI: VE0002)
(15) *Marko on farmari, hän työskentelee perheen viinitarhassa ja toivoo että myöhemmin luovuttaa sen poikalleen, joka nyt on 15 vuotias ja lapsuudestaan pitää* ***tästä*** *työstä.* (ICLFI: VE0008)

Kuviossa 6 on esitettynä pronominien määrät prosentteina koko taitotason saneista.



KUVIO 6. Pronominien prosentuaaliset osuudet kaikista taitotason saneista taitotasoilla B1 ja B2.

Kuten kuviosta 6 näkyy, pronominien osuus hieman pienenee B1-tason 13 prosentista B2-tason 11,7 prosenttiin.

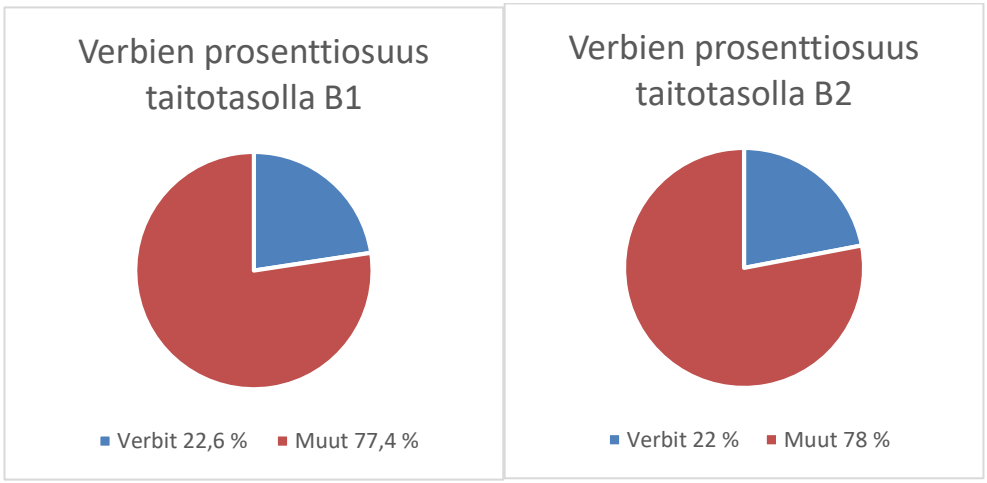
Verbien morfologisen analyysin koodi on aineistossani @MAIN V. Kaikkiaan niitä löytyy venäläisten teksteistä 23 989. Tasolla B1 6 757 ja tasolla B2 13 612. Liittomuodot, kuten esimerkiksi *on ollut*, on laskettu kahdeksi verbiksi. Verbeissä on mukana partisiipit, kuten *hämmästynyt*, ilmauksessa *oli hämmästynyt.* Tai kuten seuraavassa esimerkissä (16) *sanotut*:

(16) *Näytelmässä voi nähdä niin **sanotut** The Golden Fifties, kultaiset viisikymppiset.* (ICLFI: VE0003)

Esimerkeissä (17–18) näkyy lisää verbien käyttöä:

(17) *Hän **oli** sairaalassa ja ei **voinut** edes **nousta**, koska hänen jalkansa **oli mennyt** poikki.* (ICLFI: VE0001)
(18) ***Haluan sanoa**, että tämä näytelmä jätti huonon vaikutelman.* (ICLFI: VE0003)

Kuviossa 7 on esitettynä verbien määrät prosentteina koko taitotason saneista.



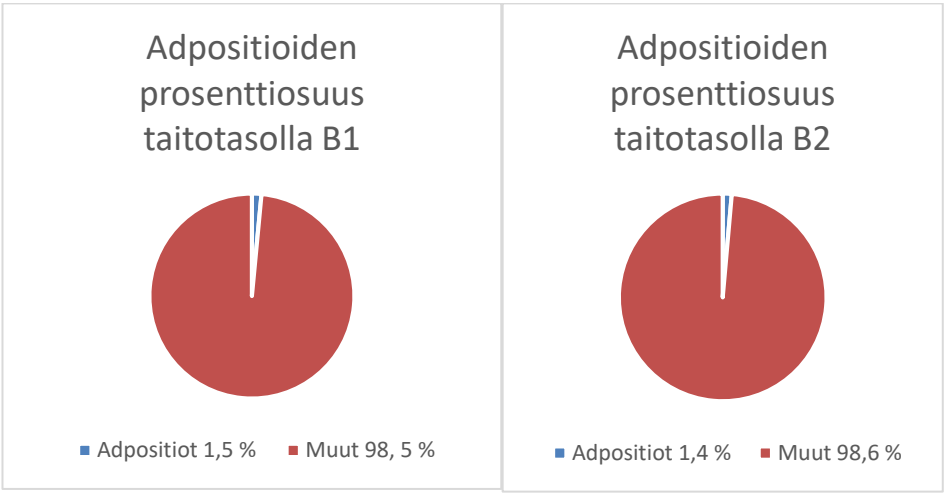
KUVIO 7. Verbien prosentuaaliset osuudet kaikista taitotason saneista taitotasoilla B1 ja B2.

Kuten kuviosta 7 näkyy, verbien osuus tason B1 saneista on 22,6 % ja tason B2 saneista 22 %.

Adpositioiden morfologisen analyysin koodit ovat @PREMARK POST postpositioille ja @PREMARK PREP prepositioille. Näillä morfologisen analyysin koodeilla osumia yhteensä 1 534, taitotasolla B1 441, B2 855. Esimerkki (19) on postposition käyttöä ja esimerkki (20) preposition käyttöä:

(19) *Voin sanoa, että loppuratkaisu on avoin, kaikki tapahtuu mediän silmien* ***vieressä****, näemme kaikki omin silmin.* (ICLFI: VE0143)
(20) *Suomen vesi on puhdasta ja suomalaisien ei tarvitse kehuttaa vettä,* ***ennen*** *juomista.* (ICLFI: VE0114)

Kuviossa 8 on esitettynä adpositioiden määrät prosentteina koko taitotason saneista.



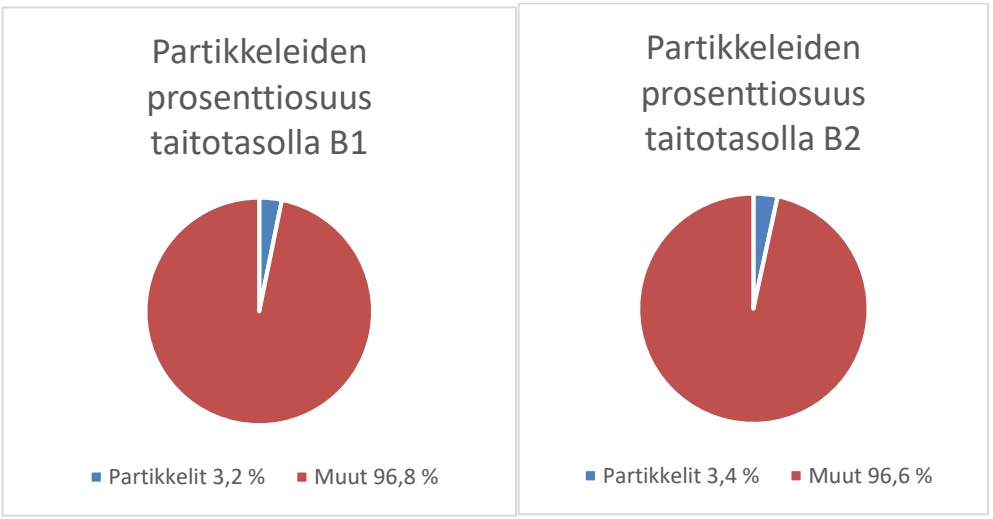
KUVIO 8. Adpositioiden prosentuaaliset osuudet kaikista taitotason saneista taitotasoilla B1 ja B2.

Kuviosta 8 näkyy, että tasolla B1 adpositioiden osuus kaikista saneista on 1,5 % ja tasolla B2 1,4 %. Tasojen välillä on vain 0,1 % ero.

Partikkelit löytyvät morfologisen analyysin koodilla @PREMARK CS ja kaikkiaan niitä löytyy aineistosta 3 560. Taitotasolta B1 952 ja taitotasolta B2 2 101. Esimerkit (21–22) kuvaavat partikkeleiden käyttöä:

(21) *Katya tulisi pöllöksi,* ***koska*** *hän on tyyni.* (ICLFI: VE0070)
(22) *Toivon,* ***että*** *se oli vain ystävällinen vitsi.* (ICLFI: VE0093)

Kuviossa 9 on esitettynä partikkeleiden määrät prosentteina koko taitotason saneista.



KUVIO 9. Partikkeleiden prosentuaaliset osuudet kaikista taitotason saneista taitotasoilla B1 ja B2.

Kuten kuviosta 9 näkyy, partikkeleiden osuus tasolla B1 on 3,2 % ja tasolla B2 se on 3,4 %. Eroa on 0,2 %.

4.3. Tulosten yhteenveto

Hakiessani aineistosta kaikkien venäjänkielisten tekstien sanemäärää, mukaan tulee myös pilkut ja pisteet. Lisäämällä hakuun komennot ”perusmuoto” ”ei ole” ”.” ja ”perusmuoto” ”ei ole” ”,” saan rajattua pilkut ja pisteet pois hakutuloksesta. Tällöin kokonaissanemääräksi tulee 108 565. Kun lisään hakuun komennon ”taitotaso CEFR:n mukaan” ”B1”, saan tulokseksi 29 962 ja taitotasolle B2 61 854. Taulukkoon 3 olen kerännyt yhteissanemäärät ja taulukkoon 4 sanaluokkien määrät.

TAULUKKO 3. Yhteissanemäärät kaikilla venäjänkielisillä ja taitotasoilla B1 ja B2.

	Aineiston kaikki venäjänkieliset	**Taitotasolla B1**	**Taitotasolla B2**
Saneet	108 565	29 962	61 854

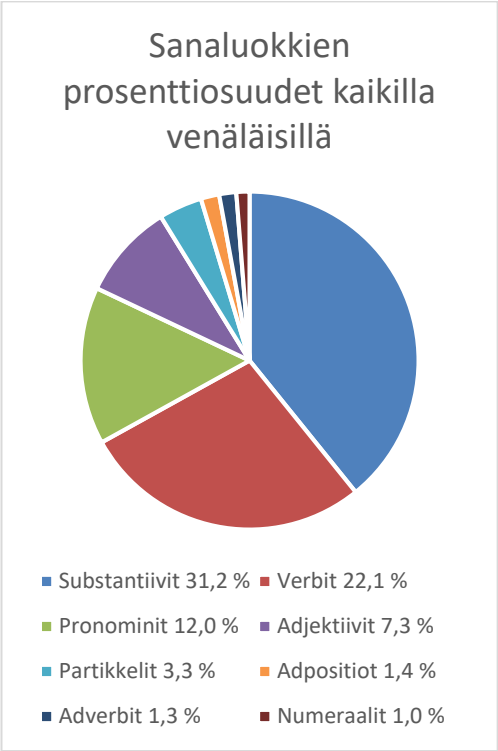
Aineistoon jää edelleen merkkejä, jotka eivät ole sanoja. Tämä johtuu siitä, että ne on korpuksessa koodattu saneiksi. Pisteet ja pilkut ovat niistä ilmeisimmät ja ne on poistettu.

Taulukon 3 yhteissanemäärissä on siis mukana myös näitä merkkejä, mutta kandidaatin tutkielman puitteissa en pystynyt selvittämään niitä kaikkia.

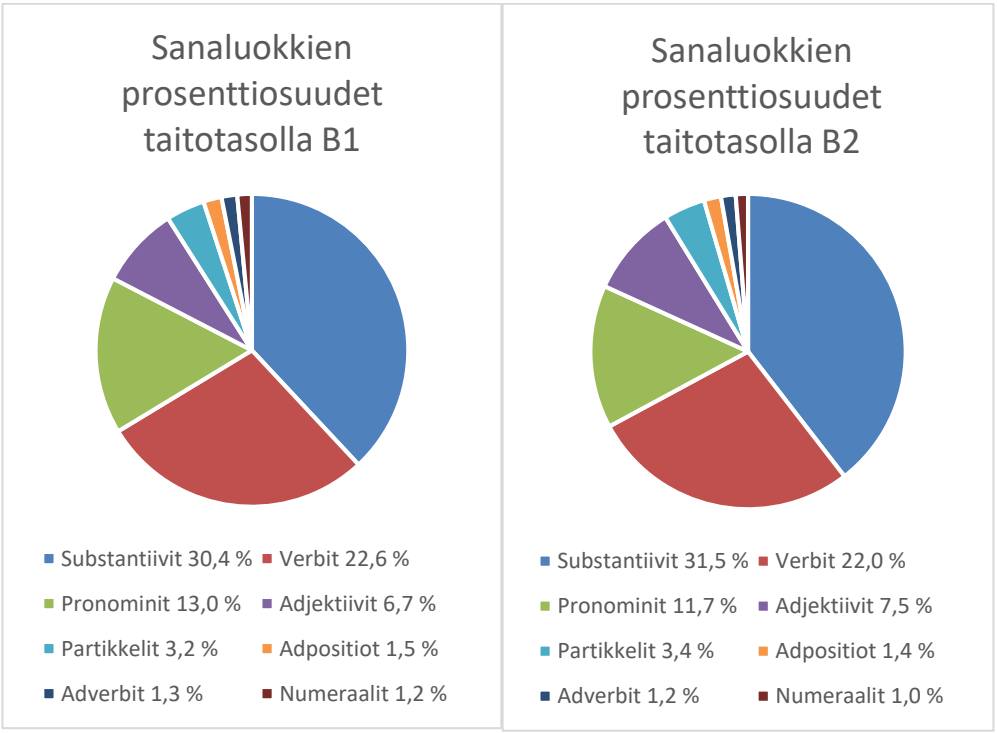
TAULUKKO 4. Sanaluokkien määrät kaikilla venäläisillä ja taitotasoilla B1 ja B2.

Sanaluokka	**Sanemäärä kaikilla venäläisillä**	**Sanemäärä taitotasolla B1**	**Sanemäärä taitotasolla B2**
Substantiivit	33 857	9 103	19 470
Verbit	23 989	6 757	13 612
Pronominit	12 932	3 899	7 216
Adjektiivit	7 928	2 012	4 639
Partikkelit	3 560	952	2 101
Adpositiot	1 534	441	8 55
Adverbit	1 387	387	760
Numeraalit	1 125	370	614

Taulukoissa 3 ja 4 esitettyjen numeroiden havainnollistamiseksi ja sanaluokkien välisten suhteiden kuvaamiseksi esitän kuviossa 10 kaikkien sanaluokkien prosenttiosuudet kaikkien venäläisten kokonaissanemäärästä ja kuviossa 11 kaikkien sanaluokkien prosenttiosuudet kummankin taitotason kokonaissanemäärästä.



KUVIO 10. Kaikkien sanaluokkien prosentuaaliset osuudet kaikista saneista kaikilla aineiston venäläisillä.



KUVIO 11. Kaikkien sanaluokkien prosentuaaliset osuudet kaikista taitotason saneista taitotasoilla B1 ja B2.

Kuvioihin 10 ja 11 on kerätty kaikkien venäläisten ja kummankin tutkittavan taitotason kaikki sanaluokat ja niitä katsoessa huomaa, että sanaluokkien osuudet taitotason kokonaissanemäärästä ovat hyvin lähellä toisiaan. Määrät jakautuvat hyvin samaan tapaan niin kaikilla venäläisillä kuin kummallakin tutkimallani taitotasolla.

Sanaluokkien osuudet kaikilla venäläisillä ovat adpositioiden, adverbien ja numeraalien osalta prosenttiyksiköiden kymmenyksiä myöten samoja jommankumman taitotason osuuden kanssa. Muiden sanaluokkien osalta kaikkien venäläisten osuudet sijoittuvat taitotasojen osuuksien väliin. Näin esimerkiksi substantiiveilla, joiden osuus kaikilla venäläisillä on 31,2 % ja B1 tasolla 30,4 % ja B2 tasolla 31,5 %. Ero ovat kuitenkin hyvin pieniä, kuten näistäkin luvuista huomaa.

Taitotasojen B1 ja B2 väliltä ei myöskään löydy huomattavia eroja. Tutkimani taitotasot ovat kumpikin B-tasoa, mikä voi selittää tulosten samankaltaisuutta. Kummallakin tasolla substantiivit ovat suurin ryhmä ja toiseksi suurin on verbit. Tämän aineiston perusteella kielitaidon kehittyessä eli siirryttäessä tasolta B1 tasolle B2 pronominien osuus pienenee. Sen sijaan substantiivien ja adjektiivien osuus kasvaa. Muiden sanaluokkien osalta muutokset ovat pieniä, prosenttiyksiköiden kymmenyksiä.

5. PÄÄTÄNTÖ

Tutkin kandidaatintutkielmassani venäjänkielisten suomenopiskelijoiden oppijansuomea. Selvitin suomen kielen sanaluokkien määrät taitotasoilla B1 ja B2 ja laskin niiden prosenttiosuudet taitotason kokonaissanemäärästä. Kontrastiivisen välikielen analyysin mukaisesti vertailin näiden kahden taitotason tuloksia. Vertasin taitotasojen tuloksia keskenään ja selvitin, kuinka määrät muuttuvat, kun kielitaito paranee ja siirrytään taitotasolla ylöspäin.

Aineistoni oli kansainvälisestä oppijansuomen korpuksesta, International Corpus of Learner Finnish (ICLFI), johon on kerätty suomen opiskelijoiden kirjoittamia tekstejä. Tutkittavat tekstit olivat opiskelijoiden opintojaan varten kirjoittamia tekstejä, jotka oli taitotasoarvioitu eurooppalaisen viitekehyksen mukaan. Tutkimani tekstit olivat taitotasoilta B1 ja B2, koska niiden tasojen sanemäärät sopivat tutkittaviksi ja vertailtaviksi keskenään.

Kaikkiaan eri sanaluokkien prosenttiosuudet olivat tutkimillani taitotasoilla hyvin lähellä toisiaan. Kummallakin tasolla substantiivien osuus oli suurin ja toiseksi suurin oli verbien osuus. Siirryttäessä tasolta B1 tasolle B2 eli kielitaidon kehittyessä substantiivien ja adjektiivien osuus kasvaa. Substantiivien osuus kasvaa 30,4 prosentista 31,5 prosenttiin ja adjektiivien osuus 6,7 prosentista 7,5 prosenttiin. Pronominien määrä taas laskee 13 prosentista 11,7 prosenttiin. Näidenkään sanaluokkien kohdalla muutokset eivät ole suuria, mutta muiden sanaluokkien kohdalla muutokset olivat prosenttiyksiköiden kymmenyksiä.

Verratessani näiden kahden taitotason sanaluokkien prosenttiosuuksia kaikkien aineiston venäläisten sanaluokkien osuuksiin, en löytänyt suuria eroja. Adpositioiden, adverbien ja numeraalien osuudet olivat prosenttiyksiköiden kymmenyksien tarkkuudella samoja kuin jommallakummalla taitotasolla. Muiden sanaluokkien osuudet sijoittuivat kaikilla venäläisillä taitotasojen osuuksien väliin. Erot olivat kuitenkin hyvin pieniä, alle prosentin suuruisia.

Jatkossa on mahdollista tutkia samalla tavalla niiden opiskelijoiden oppijansuomea, joilla on eri äidinkielet. Toisaalta venäläisten oppijansuomen voi ottaa tarkempaan tarkasteluun

ja tutkia tarkemmin eri sanaluokkien esiintymistä, nyt tarkastelun kohteena oli vain sanaluokkien määrät. Mahdollinen tutkimuksen kohde olisi myös eri sanaluokkien osuuksien tutkiminen esimerkiksi A- ja B-tasoilla tai B- ja C-tasoilla. Muuttuisivatko osuuksien suhteet ja jos muuttuisivat niin miten. Tässä tutkimuksessa käsiteltiin vain Venäjällä asuvien, venäjää äidinkielenään puhuvien suomen oppijoiden suomea. Jatkossa on mahdollista vertailla Virossa ja Venäjällä asuvien, venäjää äidinkielenään puhuvien suomen oppijoiden suomen kieltä. Onko esimerkiksi ympäristöllä, sillä mitä kieltä ympärillä puhutaan, vaikutusta opeteltavaan suomeen ja jos on niin millainen vaikutus.

LÄHTEET

EVK = EUROOPPALAINEN VIITEKEHYS. Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys 2003. Porvoo: WSOY.

GRANGER, SYLVIANE 2002: A bird's eye view of learner corpus research. – Sylviane Granger, Heidrun Jung & Stephanie Petch-Tyson (toim.), *Computer learner corpora. Second language acquisition and foreign language teachings* 3−33. Amsterdam: John Benjamins Publishing Company.

ICLFI-MANUAALI. https://www.oulu.fi/sites/default/files/content/ohjeet_korpuksen_kayttajalle.docx (10.10.2019).

HUTTU-HILTUNEN, MARIA 2017: *Joka-, mikä-, ja kuka-relatiivikonstruktioiden virheet virolaisten suomenoppijoiden kirjoitelmissa.* Pro gradu -tutkielma. Oulun yliopiston suomen kielen oppiaine.

ISK = HAKULINEN, AULI – VILKUNA, MARIA – KORHONEN, RIITTA – KOIVISTO, VESA – HEINONEN, TARJA RIITTA – ALHO, IRJA 2004: *Iso suomen kielioppi.* SKST 950. Helsinki: SKS.

IVASKA, ILMARI 2015: *Edistyneen oppijan suomen konstruktiopiirteitä korpusvetoisesti: avainrakenneanalyysi.* Annales Universitatis Turkuensis C 409. Turku: Turun yliopisto.

JANTUNEN, JARMO HARRI 2004: *Synonymia ja käännössuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin.* University of Joensuu Publications in the Humanities 35. Joensuu: Joensuun yliopisto.

JANTUNEN, JARMO HARRI – BRUNNI, SISKO – LEHTO, LIISA-MARIA – AIRAKSINEN, VALTTERI 2014: Oppijankieliaineistojen annotointi – esimerkkinä ICLFI:n annotoinnin prosessit, ongelmat ja ratkaisut. *AFinLA-E: Soveltavan Kielitieteen Tutkimuksia* (7) s. 60–80. https://journal.fi/afinla/article/view/48160 (10.10.2019).

JANTUNEN, JARMO HARRI – BRUNNI, SISKO – OULUN YLIOPISTO 2013: *Kansainvälinen oppijansuomen korpus* [tekstikorpus]. Kielipankki. http://urn.fi/urn:nbn:fi:lb-20140730163 (10.10.2019).

JANTUNEN, JARMO HARRI – PIRKOLA, SILJA 2015: Oppijansuomen sähköiset tutkimusaineistot: nykytilanne. – *Virittäjä* 119 (1) s. 88–103.

KANSALAISUUSLAKI 16.5.2003/359. http://www.finlex.fi/fi/laki/ajantasa/2004/20030359 (3.12.2019)

KAJANDER, MIKKO 2013. *Suomen eksistentiaalilause toisen kielen oppimisen polulla.* Jyväskylä Studies in Humanities 220. Jyväskylä: Jyväskylän yliopisto.

KUULUVAINEN, HELENA 2015: *Fraseologiset virheet Kansainvälisessä oppijansuomen korpuksessa.* Pro gradu -tutkielma. Oulun yliopiston suomen kielen oppiaine.

LATOMAA, SIRKKU 1996: Matkalla uuteen kieleen. – Ruuska H. & Tuomi, S-M (toim.) *Moneja baareja: Tiellä toimivaan kaksikielisyyteen.* Äidinkielen opettajain liiton vuosikirja 1996 s. 97–106. Helsinki: ÄOL.

LATOMAA, SIRKKU 1993: Mitä hyötyä on oppijoiden kielitaustan tuntemisesta? – Eija Aalto & Minna Suni (toim.), *Kohdekielenä suomi. Näkökulmia opetukseen* s .9–31. Jyväskylä: Korkeakoulujen kielikeskus.

LIITI, ANNIKA 2018: *Ruotsinkielisten suomenoppijoiden morfosyntaktisten verbimuotojen virhekäytöt.* Pro gradu -tutkielma. Oulun yliopiston suomen kielen oppiaine.

LOUNELA, MIKKO – HEIKKINEN, VESA 2012: Korpus. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiililä & Mikko Lounela (toim.) *Genreanalyysi: tekstilajitutkimuksen käsikirja* s. 120–127. Kotimaisten kielten keskuksen julkaisuja 169. Helsinki: Gaudeamus.

MUSTONEN, SANNA 2015. *Käytössä kehittyvä kieli. Paikat ja tilat suomi toisena kielenä -oppijoiden teksteissä.* Jyväskylä Studies in Humanities 255. Jyväskylä: Jyväskylän yliopisto.

NISSILÄ, LEENA 2011. *Viron kielen vaikutus suomen kielen verbien ja niiden rektioiden oppimiseen.* Acta Universitatis Ouluensis, B Humaniora 99. Oulu: Oulun yliopisto.

PIRI, OLLI-JUHANI 2017: *Morfosyntaktiset objektivirheet oppijansuomessa. Korpuspohjainen kuvaus suomenoppijoiden kielitaidon tarkkuuden kehityksestä.* Pro gradu -tutkielma. Oulun yliopiston suomen kielen oppiaine.

SPOELMAN, MARIANNE 2013: *Prior linguistic knowledge matters. The use of the partitive case in Finnish learner language.* Acta Universitatis Ouluensis, B Humaniora 111. Oulu: Oulun yliopisto.

TOGNINI-BONELLI, ELENA 2001: *Corpus Linguistics at Work.* Studies in Corpus Linguistics 6. Amsterdam: John Benjamins Publishing Company.

VARRIO, HANNA 2014: *Sanoa-verbi fraseologisena yksikkönä oppijansuomessa ja natiivikielessä. Verbin ydin ja täydennysympäristö.* Pro gradu -tutkielma. Oulun yliopiston suomen kielen oppiaine.