

**Sanojen kompleksisuuden vaikutus astevaihtelun  
oppimiseen suomi toisena kielenä -oppijoiden  
teksteissä**

Melina Vetikko

Pro gradu -tutkielma

Kielen oppimisen ja opettamisen tutkinto-ohjelma, suomen kieli

Kieli- ja käännöstieteiden laitos

Humanistinen tiedekunta

Turun yliopisto

Huhtikuu 2024

## Pro gradu -tutkielma

### Kielen oppimisen ja opettamisen tutkinto-ohjelma, suomen kieli

Melina Vetikko

### Sanojen kompleksisuuden vaikutus astevaihtelun oppimiseen suomi toisena kielenä - oppijoiden teksteissä

Sivumäärä: 83 s.

Pro gradu -tutkielmassani tutkin sanojen kompleksisuuden vaikutusta astevaihtelun oppimiseen. Tutkin suomi toisena kielenä oppijoiden käyttämiä *k:n* astevaihtelusanoja sekä kvalitatiivisessa että kvantitatiivisessa vaihtelussa. Tutkin käytettyjä muotoja erikseen tarkastellen eri vaihteluparien, eri sijamuotojen, tavumäärän, sanatyypin sekä sanan syntaktisen funktion vaikutusta muotojen norminmukaisuuteen. Käytän tutkielmassani Jyväskylän yliopiston *Cefling*-hankkeessa kehitettyä DEMfad-mallia, jossa kielenilmiötä tutkitaan sen ensiesiintymisestä hallintaan tarkastelemalla frekvenssiä, tarkkuutta ja jakaumaa. Tarkastelen käytettyjen muotojen muutosta Eurooppalaisen viitekehyksen taitotasolla.

Poimin aineiston Kansainvälisestä oppijansuomen korpukselta (ICLFI) tietyin hakutermein. Poimin sanat korpukselta satunnaisesti niin, että jokaiselta taitotasolta otantaan valikoitui sata kvalitatiivisen ja sata kvantitatiivisen astevaihtelun tapausia pois lukien muutama taitotaso, joilla ei ollut tarpeeksi osumia. Tämän jälkeen sijoitin aineiston taulukkolaskelmaohjelmaan, johon analysoin jokaisen sanan tarkasteltavien muuttujien valossa. Tutkielmani analyysi on pääosin määrällistä, mutta tarkastelen astevaihtelun kannalta virheellisiä muotoja myös yksityiskohtaisemmin.

Kvantitatiivinen vaihtelu on tutkielman mukaan oppijoille hieman haastavampi kuin kvalitatiivinen vaihtelu, ja keskimäärin astevaihtelu on paremmin hallinnassa ylemmillä taitotasolla. Taitotasolla B2 astevaihtelun käyttö on jo pääosin täysin norminmukaista, ja kokonaisuudessaan astevaihtelu on hallinnassa jokaisella taitotasolla. *k:n* astevaihtelua esiintyy ensimmäisen kerran jo aivan ensimmäisellä taitotasolla. Vaihtelupareista eniten virheellisiä muotoja tuotti *nk : ng* -vaihtelupari ja sijamuodoista tarkastelutavasta riippuen joko translatiivi tai genetiivi. Eniten virheitä oli kaksi- ja kolmetavuisissa sanoissa, joihin lisätään päätte, joka on joko muotoa CCV tai C. Vaikein sanatyypit astevaihtelun kannalta on PANKKI, ja eniten virheellisiä muotoja oli sellaisissa sanoissa, jotka ovat adverbiaaleja. Astevaihtelusanojen frekvenssi kehittyi tasaisesti ensimmäisillä taitotasolla, minkä jälkeen se saavuttaa huippunsa taitotasolla B1 ja B2. Tämän jälkeen frekvenssi laskee melkein lähtötasolle, josta se nousee vielä viimeisellä taitotasolla hieman. Jakauma vaihtelee jokaisen tarkasteltavan muuttujan osalta melko paljon taitotasoin.

**Avainsanat:** astevaihtelu, suomi toisena kielenä, kielen oppiminen, kompleksisuus

# Sisällysluettelo

<b>1</b>	<b>Johdanto</b>	<b>5</b>
<b>2</b>	<b>Suomen sanojen kompleksisuus ja astevaihtelu</b>	<b>10</b>
2.1	Suomi opittavana kielenä	10
2.2	Kielen kompleksisuuden tutkimus	11
2.3	Astevaihtelu	13
2.4	Astevaihtelu suomi toisena kielenä -oppijoilla	14
2.5	Eurooppalainen viitekehys	16
<b>3</b>	<b>Aineisto ja metodit</b>	<b>18</b>
3.1	ICLFI-korpus	18
3.1.1	Yleistä	18
3.1.2	Aineiston poiminta korpuksesta	19
3.2	Aineiston sanoista tarkasteltavat tekijät	20
3.2.1	Määrällinen ja laadullinen vaihtelu sekä vaihteluparit	20
3.2.2	Sijamuodot	21
3.2.3	Sanan pituus tavuina	21
3.2.4	Sanatyypit	22
3.2.5	Vartalon muutokset	23
3.2.6	Syntaktinen funktio	24
3.3	DEMfad-malli	24
<b>4</b>	<b>Astevaihtelusanojen kompleksisuuden vaikutus niiden oppimiseen</b>	<b>28</b>
4.1	Kvalitatiivisen vaihtelun sanat	28
4.1.1	Frekvenssi	28
4.1.2	Vaihteluparien jakauma ja tarkkuus	30
4.1.3	Sijamuotojen jakauma ja tarkkuus	34
4.1.4	Sanojen pituuden jakauma ja tarkkuus	41
4.1.5	Sanatyyppien jakauma ja tarkkuus	44
4.1.6	Sanojen syntaktisen funktion jakauma ja tarkkuus	48
4.2	Kvantitatiivisen vaihtelun sanat	51
4.2.1	Frekvenssi	51
4.2.2	Vaihteluparin jakauma ja tarkkuus	52
4.2.3	Sijamuotojen jakauma ja tarkkuus	54
4.2.4	Sanojen pituuden jakauma ja tarkkuus	61

4.2.5	Sanatyyppien jakauma ja tarkkuus	66
4.2.6	Sanojen syntaktisen funktion jakauma ja tarkkuus	69
<b>5</b>	<b>Lopuksi</b>	<b>73</b>
5.1	Yhteenveto	73
5.2	Päätelmät	76
	<b>Lähteet</b>	<b>79</b>
	<b>Aineslähteet</b>	<b>79</b>
	<b>Muut lähteet</b>	<b>79</b>

# 1 Johdanto

Suomen kieli on agglutinoiva kieli, jonka määritelmänä on, että sellaisessa kielessä sanojen päätteitä liitetään sanavartaloihin muuttamatta vartalosta mitään (Tieteen termipankki, Kielitiede: agglutinoiva kieli, 2022, sähköinen lähde). Suomen kielessä nomineihin voi liittää monenlaisia päätteitä, jotka ilmaisevat johdinten lisäksi lukua, sijaa ja omistajaa. Nomineihin voi liittää myös liitepartikkeleja. (Karlsson 2009, 35.) Suomen kieli on kuitenkin siinä mielessä poikkeava puhtaasti agglutinoivista kielistä, että tällaisten päätteiden lisääminen voi, ja usein laukaiseekin, äännevaihteluita sanan vartalossa. Tällöin sanan vartalot voivat olla erinäköisiä riippuen siihen liitettävistä päätteistä. Tästä syystä ei riitä, että puhuja osaa muodostaa sanasta vartalon ja tietää siihen liitettävän päätteen, koska monet sanat eivät taivu suomen kielessä näin yksinkertaisesti. Äännevaihteluita on monenlaisia, mutta tärkein niistä on astevaihtelu. Astevaihtelu koskee klusiileja *p*, *t* ja *k*. (Karlsson 2009, 45). Astevaihtelun oppiminen on kielenoppijalle tärkeää, koska se on suhteellisen yleinen ilmiö suomen kielessä. Taajuussanaston tuhannen yleisimmän sanan joukosta vajaa kolmannes sana taipuu niin, että päätte laukaisee astevaihtelun. (VISK § 41.)

Suomen kielen opetuksen historia yltää lähes 200 vuoden taakse, kun suomenruotsalaisille alettiin opettaa suomea. Suomea toisena kielenä alettiin opettaa 1950-luvulla Helsingin yliopistoon tulleille ulkomaalaisille opettajille, ja opetusta on jatkettu yhä kasvavissa määrin esimerkiksi ulkomailta tulleille opiskelijoille. (Mátyás 1999, 19.) Suomi toisena kielenä -alaa on tutkittu vuodesta 1967, ja vielä vuonna 1996 alaa on luonnehdittu nuoreksi (Aalto–Latomaa–Sunni 1996, 6). Marjut Vehkanen (2015) on väitöskirjassaan tutkinut suomi toisena ja vieraana kielenä -oppikirjoja vuosina 1866–1953. Väitöskirjasta käy ilmi, että astevaihtelua on esitelty jokaisessa kielioppia sisältävässä oppikirjassa, joita hän tarkasteli (Vehkanen 2015, 90). Astevaihtelu on siis ollut osana suomen kielen opettamista jo hyvin pitkän aikaa, joka kertoo siitä, että astevaihtelu on hyvin keskeinen ilmiö suomen kielessä. Siihen tulee kiinnittää huomiota erityisesti silloin, kun suomea opiskelee vieraana kielenä.

Tässä tutkielmassa tarkastelen sanojen kompleksisuuden vaikutusta astevaihtelun oppimiseen suomi toisena kielenä -oppijoilla. Tutkielmaani motivoi aito kiinnostus sellaisen kielenpiirteen oppimiseen, jolla ei periaatteessa ole mitään kieliopillista merkitystä, ja joka silti on suomen kielessä hyvin yleinen ilmiö. Aivan merkityksetön astevaihtelu ei kuitenkaan ole, koska se voi synnyttää minimipareja, jolloin sen käyttäminen tai käyttämättömyys luo merkityseroja (vrt. *tavata* : *tavaan*; *tavata* : *tapaan*). Olen aiemmin tutkinut astevaihtelun

oppimista kandidaatintutkielmassani (Vetikko 2021), jossa tarkastelin sitä, millä Eurooppalaisen viitekehyksen taitotasolla suomenoppijat oppivat  $k$ :n astevaihtelun sellaisissa sanoissa, jotka ovat genetiivissä. Tässä tutkielmassa tutkin  $k$ :n astevaihtelun oppimista kaikissa sijamuodoissa ja siihen vaikuttavia muita tekijöitä. Tarkasteltavat tekijät ovat määrällinen ja laadullinen vaihtelu, vaihteluparit, sijamuodot, sanatyypit, sanan pituus tavuina sekä syntaktinen funktio.

Suomeen on viime vuosina tullut ennätysmäärä maahanmuuttajia (Tilastokeskus, Maahanmuutto ennätyskellisen korkealla viime vuonna, 2024, sähköinen lähde). Jos maahan muuttanut henkilö haluaa Suomen kansalaisuuden, hänellä on erilaisia edellytyksiä, joista yksi on se, että hän osaa vähintään tyydyttävällä tasolla suullista ja kirjallista suomea, suullista ja kirjallista ruotsia, suomalaista viittomakieltä tai suomenruotsalaista viittomakieltä (Maahanmuuttovirasto, Luvat ja kansalaisuus, 2022, sähköinen lähde). Vaikka henkilö ei haluaisikaan Suomen kansalaisuutta, kohdemaan kielitaidolla on silti suuri merkitys. Esimerkiksi yksi työllistymiseen vaikuttava tekijä on kohdemaan kielen osaaminen. Tätä näkemystä tukevat sekä kielikoulutuspoliittiset linjaukset että tutkimukset. (Pöyhönen–Tarnanen 2011, 139.) Kieltä voi toki oppia myös arkielämässä, mutta on olemassa tutkimustuloksia, joiden mukaan kieliopin opettaminen tehostaa ja nopeuttaa uuden kielen oppimista verrattuna naturalistiseen oppimiseen (Jaakkola 1997, 77). Suomen kielen opetuksen tarve on siis suuri, ja sen vuoksi sitä on myös tarpeellista tutkia.

Suomen kieltä on totuttuun tapaan opetettu perinteisen kielenopetuksen tradition mukaisesti, ja siinä on keskitytty kielioppisääntöjen välittämiseen ja siihen, että muotoja osattaisiin muodostaa virheittä. Koska morfologia on niin keskeinen asia suomen kielessä, myös fennistiikassa siihen on kiinnitetty paljon huomiota. (Mátyás 1999, 25.) Leila White mainitsee *Suomen kielioppia ulkomaalaisille* -kirjansa esipuheessa, että on halunnut tehdä morfologian osuudesta suuren, koska taivutusoppi on pohja, jolle kaikki muu suomen kielen osaaminen rakentuu (2019, 5). Monet suomea vieraana kielenä opettaneet ovat sitä mieltä, että kaikista morfofonologisista vaihteluista astevaihtelu aiheuttaa oppijoille eniten hankaluuksia. Tämä ei sinänsä ole yllättävää, koska astevaihtelu on niin yleinen ilmiö suomen kielessä. Astevaihtelusanoja tulee esiin lähes joka lauseessa, minkä vuoksi sen opettaminen on tärkeä asia. (Karlsson 1981, 100.)

Oppijan käyttämän kielen ja kirjoittaman tekstin kognitiivista ja kielellistä monimutkaisuutta kutsutaan kompleksisuudeksi. Kompleksisuutta voidaan mitata määrällisillä mittareilla.

Oppijan ilmauksista voidaan mitata esimerkiksi keskimääräinen sana- ja morfeemimäärä tai lauseiden määrä tai niissä käytetyt rinnastukset ja alistukset. On kehitetty myös monia muita tapoja arvioida kielen kompleksisuutta ja kompleksisuuden käsitteestä keskustellaankin jatkuvasti esimerkiksi kielen oppimista tutkivien tutkijoiden keskuudessa. (Alanen – Martin 2011, 39–40.) Toisen kielen oppimista paljon tutkineet Bulté ja Housen kirjoittavat artikkelissaan (2012), että yksinkertaisimmillaan kompleksisuus on ilmiön tai kokonaisuuden ominaisuudet tai laatu sen suhteen, kuinka monesta ja millaisesta komponentista nämä ilmiöt tai kokonaisuudet rakentuvat sekä näiden keskinäisten suhteiden määrä ja luonne.

Toisen kielen oppimista on tutkittu monista näkökulmista. Yleisesti kielenopetuksessa virheisiin on kiinnitetty paljon huomiota. Virheitä pyritään yleensä välttämään, mutta jo 1980-luvulla huomattiin, että virheet oikeastaan auttavat oppijaa eteenpäin ja toimivat hyvänä lähteenä oppimiselle. Virheanalyysiin keskittyvä tutkimus oli suosituinta 1970-luvulla, mutta se tuli suomalais-ugrilaisten kielten tutkimukseen vasta myöhemmin. (Suni 2012, 410–411.) Oppijan tekemät virheet ovat kuitenkin tärkeitä tunnistaa, koska virheettömästä muodosta ei pysty tunnistamaan muodostamisprosessia. Kun oppija tekee virheen, tuottamisprosessi voi olla näkyvässä, ja virheen lähde voidaan tunnistaa. Virheiden tarkastelulla voidaan erityisesti tunnistaa kielten välinen vaikutus. (Martin 1995, 264.)

Virheanalyysin lisäksi 1970-luvulla alettiin tutkia välikieltä. Välikielen konseptissa keskitytään siihen, miten oppijan henkilökohtainen kielivarieteetti pikkuhiljaa muuttuu kohti kielimuotoa, jota tavoitellaan. Virheanalyysiä pystytään kuitenkin nykyisin hyödyntämään yksityiskohtaisemmin, mikä voi johtaa yhä edelleen relevantteihin tutkimustuloksiin, jotka ovat hyödyllisiä. (Suni 2012, 410–411.) Myös vertailevaa analyysiä on tehty esimerkiksi tutkimalla kieltenvälisiä vaikutuksia. Vertailevat lähtökohdat ovat olleet suosittuja suomalais-ugrilaisten kielten tutkimuksissa, vaikka välttämättä pääpainona ei olekaan ollut tutkia tätä toisen kielen oppimisen näkökulmasta. (Suni 2012, 412.) Myös muotoja ja rakenteita on tutkittu paljon. Kieliopillisten morfeemien oppimista on tutkittu etenkin englannin kielen näkökulmasta, mikä tietysti eroaa paljon suomalais-ugrilaista kielistä, koska morfeemeja on enemmän ja niiden järjestykset ovat monimutkaisia. Suomalais-ugrilaisten kielten oppimisessa toisena kielenä on tutkittu paljon morfologiaa, mikä ei edellä mainittujen syiden takia ole yllättävää. (Suni 2012, 417.) Toisen kielen oppimista on tutkittu myös vuorovaikutuksen näkökulmasta. Tutkimusta on tehty esimerkiksi niin, että on keskitytty siihen, miten vuorovaikutuksessa olevat henkilöt korjaavat omia kielen muotojaan niin, että informaatio välittyy ja vuorovaikutusta voi jatkaa. (Suni 2012, 420).

Kielen oppiminen on hyvin moniulotteinen ilmiö, eikä siitä sen vuoksi pystytä luomaan teoriaa tai kuvausta, joka kattaisi kaiken, ja sitä onkin sen takia tutkittu hyvin monesta näkökulmasta. Onkin olemassa monenlaisia kielen oppimisen teorioita, jotka kukin pyrkivät selittämään eri ilmiöitä eri tavoin. Joskus nämä teorit täydentävät toisiaan, ja toisinaan uusi teoria voi syrjäyttää vanhan vallalla olleen teorian. (Lintunen–Pietilä 2014, 5.) Toisen kielen oppimisen tutkimuksessa yritetään ottaa selvää siitä, miten ihminen oppii kieliä sen jälkeen, kun hän on lapsena jo omaksunut yhden tai useamman kielen. Oppiminen on tiedostettua toimintaa, ja sen seurauksena kielenoppija oppii säännöt, joiden avulla hän osaa muodostaa oikeita muotoja ja kysyttäessä myös perustella ne. (Lintunen–Pietilä 2014, 9–10.) Toisen kielen oppimisen tutkimus voi auttaa toisen kielen oppimisen teorioita esimerkiksi kehittämällä, optimoimalla ja näyttämällä toteen niitä. Käytännössä toisen kielen oppimisen tutkimus voi auttaa pedagogoja luomaan opetukseen materiaaleja, testejä ja metodeja selvittämällä tärkeää informaatiota, joiden avulla näitä voidaan kehittää. (Winke–Brunfaut 2020, 1.)

Tutkimuskysymykseni ovat:

1. Miten astevaihtelun käytön tarkkuus korreloi oppijan kielitaitotason kanssa?
2. Miten astevaihtelun alaisten sanojen osuus muuttuu taitotasojen välillä, ja mikä niiden osuus on kaikista korpuksen substantiiveista?
3. Miten käytettyjen astevaihtelun alaisten sanojen rakenne ja syntaktinen funktio vaikuttavat eri muotoihin eri taitotasoilla?
4. Millaisia eroja on eri vaihteluparien oppimisen välillä?

Tutkin astevaihtelun kehitystä DEMfad-metodin avulla, jonka lisäksi tarkastelen käytettyjen muotoja tarkemmin. DEMfad-metodin kirjaimet edustavat kukin omaa osiotaan. D tarkoittaa tutkittavaa kielenilmiötä, E ilmiön ensiesiintymistä ja M sitä tasoa, jolla ilmiö on hallinnassa, eli 80 % esiintymistä on kohdekielen mukaisia. Kirjain f tarkoittaa frekvenssiä, a tarkoittaa norminmukaisuutta ja d jakaumaa. Tutkielman aineistona on Kansainvälinen oppijansuomen korpus (ICLFI), josta poimin oppijoiden käyttämiä sanoja tietyin kriteerein. Esittelen metodia ja korpusta enemmän luvussa 3.

Etenen tutkielmassani niin, että ensin luvussa 2 keskityn esittelemään taustaa tutkimukselleni. Tässä luvussa kerron suomen kielen erityispiirteistä, kielen kompleksisuuden tutkimuksesta

sekä astevaihtelusta yleisesti ja erityisesti suomi toisena kielenä -oppijoilla. Luvussa 3 esittelen aineistoa ja metodia, sekä tarkemmin tarkastelun alaisia tekijöitä. Luvussa 4 esittelen määrällisen ja laadullisen analyysin muuttujakohtaisissa luvuissa erikseen sekä kvalitatiivisen että kvantitatiivisen vaihtelun alaisten sanojen osalta. Luvussa 5 kokoan analyysiä yhteen ja kirjoitan tutkimuksen tuloksista.

## 2 Suomen sanojen kompleksisuus ja astevaihtelu

### 2.1 Suomi opittavana kielenä

Monet suomalaiset ja suomen kielen kanssa tekemisissä olevat ulkomaalaiset ajattelevat, että suomen kieltä on vaikea oppia (Miestamo 2006, 1). Euroopan valtakielisiin verrattuna tämä voikin tuntua siltä. Suomen kieli kuuluu uralilaiseen kielikuntaan, ja siksi se poikkeaa suurimmasta osasta muista Euroopan kielistä, koska ne kuuluvat indoeurooppalaiseen kielikuntaan. Tavallisesti kielet ryhmitellään niin, että uralilaiset kielet ovat agglutinoivia ja indoeurooppalaiset kielet flekteeraavia kieliä. (Dahl 2008, 545.) Kun kieliä tarkastellaan laajemmin, voidaan kuitenkin todeta, että useimmat kielet, suomi mukaan lukien, yhdistävät sekä flektiota että agglutinaatiota (mt., 547). Esimerkiksi jos verrataan suomen ja muiden eurooppalaisten kielten finiittistä verbimorfologiaa, suomen kieli on vähemmän kompleksinen etenkin, jos sitä verrataan romaanisiin kieliin. Suomen kielen verbimorfologia kuitenkin vaikuttaa kompleksisemmalta, koska suomen verbit taipuvat myös persoonassa. (mt., 549.) Dahlin näkemystä on kuitenkin myös kritisoitu. Esimerkiksi Anneli Pajunen on artikkelissaan pyrkinyt osoittamaan viitatessaan Dahlin artikkeliin, että suomen sanojen synteettisyyttä on mitattu virheellisesti tai sekoittamalla eri laskutapoja (Pajunen 2010, 481–482). Synteettisyys kielessä tarkoittaa sitä, että kieli on sitä synteettisempi, mitä enemmän kielen sanoissa on morfeemeja keskimäärin. Euroopan indoeurooppalaisiin kieliin verrattuna suomi on hyvinkin synteettinen, mutta maailmanlaajuisesti tarkasteltuna suomi ei ole erityisen synteettinen kieli. (Miestamo 2006, 2.)

Muotoja on siis joka tapauksessa paljon, jos vertaa esimerkiksi johonkin indoeurooppalaiseen kieleen, mikä voi luoda suomenoppijalle mielikuvan taivutusparadigmojen vaikeudesta. Jos kuitenkin tarkastellaan morfologian muita osia, esimerkiksi suomen sijajärjestelmää, suomen kieli on todellakin morfologisesti kompleksisempi kuin useimmat muut kielet. Morfologia on kompleksista, koska nominit ja sen lisäksi myös infinitiivit taipuvat sijamuodoissa, ja tämä antaa suomen kielestä agglutinoivan kuvan. (Dahl 2008, 549.) Karlsson (2009) mainitsee muutamia tekijöitä, jotka saattavat aiheuttaa suomea toisena kielenä opiskeleville vaikeuksia. Hän nimeää näiksi muutamia kielenilmiöitä, jotka ovat ääntämyksen osalta äänneiden pituus, koska pitkät ja lyhyet vokaalit ja konsonantit voivat luoda merkityseroja, sanaston osalta hän mainitsee suomen perussanaston, joka poikkeaa indoeurooppalaisesta perussanastosta, ja morfologian osalta hän mainitsee sanojen vartalot ja niissä tapahtuvat morfofonologiset vaihtelut.

Varovasti laskettuna suomen kielen nomineista voi muodostaa noin 2 100 erilaista sananmuotoa ja verbeillä infinitiivit mukaan laskettuna noin 12 000 sananmuotoa. Mukaan ei ole laskettu johtimia, eikä syntaktisia rajoituksia ole verbien osalta huomioitu. Näinkin suureen nominimuotojen määrään päästään, jos mukaan lasketaan, että lukuja on kaksi, sijoja on 13 (pois lukien instruktiivi ja akkusatiivi), possessiivisuffikseja kuusi, liitepartikkeleja tai niistä muodostettuja jonoja 12. Tähän lisätään vielä vaihtoehtoiset 3. persoonan possessiivisuffiksit, jotka kasvattavat määrää kymmenellä vaihtoehdolla. Tällä laskutavalla on tosin mukana myös homonymiset muodot. (Karlsson 1983, 356–357.) Joka tapauksessa mahdollisia muotoja on silti valtava määrä. Astevaihtelun laukaisee kuitenkin vain suoraan sanavartaloon liitettävät päätteet, jolloin esimerkiksi liitepartikkelin jonon pituus ei vaikuta sanavartalon morfofonologisiin vaihteluihin. Sen lisäksi, että suomenoppijan tarvetta osata käyttää kieltä pidetään tärkeänä, myös monimutkaisen morfologian oppimisen kannalta kieliopin läpikotaista opettamista pidetään tärkeänä. Jos kielenoppija oppii väärät taivutusparadigmat tai hän muodostaa muutenkin virheellisiä muotoja, näiden kitkeminen myöhemmin voi olla haastavaa, minkä takia olisi tärkeää saada oppijat oppimaan heti oikeat muodot. (Aalto 1992, 7.)

## 2.2 Kielen kompleksisuuden tutkimus

Suomen kieltä siis pidetään yleisesti vaikeana, mutta asia ei ole näin yksinkertainen. Pitkään suomea opiskeli vieraana kielenä lähinnä äidinkielenään indoeurooppalaisia kieliä puhuvat ihmiset, jolloin suomea luonnollisesti verrattiin näihin kieliin. Kyse ei kuitenkaan ole kielen vaikeudesta, vaan siitä, miten ja kuinka paljon opittava kieli eroaa oppijan äidinkielestä. Kielen kompleksisuus liittyy kielen vaikeuteen, ja kompleksisuutta voi lähestyä monesta näkökulmasta. Tietyn näkökulman mukaan kielen vaikeus nähdään kompleksisuuteena. Kielitieteessä on pitkään ajateltu, että kaikki kielet ovat kokonaisuudessaan yhtä monimutkaisia niin, että jos jokin kielen osuus on yksinkertainen, toisen osuuden on oltava monimutkainen. Esimerkiksi jos kieli on morfologisesti kompleksinen, se on vähemmän kompleksinen syntaksin kannalta. Tätä ajatusta on kuitenkin alettu haastaa. Kielissä on ominaisuuksia, jotka ovat äidinkielestä riippumatta hankalia, ja tällainen haastavaksi mielletty ominaisuus on esimerkiksi sellainen taivutusjärjestelmä, jossa on paljon äännevaihteluita. (Miestamo 2006, 7–8.)

Sujuvuus (*fluency*), tarkkuus (*accuracy*) ja kompleksisuus (*complexity*) ovat ulottuvuuksia, joiden avulla tarkastellaan toisen kielen tuottamista, taitoja ja kehittymistä. Tämä on

tutkimusala, jolla pyritään vastaamaan sellaisiin kysymyksiin, kuten mikä tekee kielenkäyttäjistä pätevän ja miten toisen kielen pätevyyttä voidaan mitata kaikkein asianmukaisimmin. (Housen–Kuiken–Vedder 2012, 1.) Monet toisen kielen oppimisen tutkijat ovat sitä mieltä, että pätevyys toisessa kielessä ei ole yksi yhtenäinen asia, vaan ennemminkin se koostuu monesta osatekijästä. Olennaisimmat osatekijät voidaan tavoittaa kompleksisuuden (*complexity*), tarkkuuden (*accuracy*) ja sujuvuuden (*fluency*) käsitteillä, ja tämä voidaan lyhentää muotoon CAF. (mt., 1.)

CAF-tutkimukseen liittyy myös haasteita. Yksi haaste on esimerkiksi se, miten tutkijat määrittelevät nämä eri osatekijät. Monissa tutkimuksissa, joissa toisen kielen oppimista lähestytään kompleksisuuden, tarkkuuden ja sujuvuuden näkökulmasta, ei määritellä eksplisiittisesti, mitä näillä termeillä tarkoitetaan, tai ne määritellään suurpiirteisesti tai tarkoin kvantitatiivisin mittarein. Tämän takia kompleksisuuden, tarkkuuden ja sujuvuuden termejä käytetään eri merkityksin eri tutkimuksissa, eivätkä tutkimukset näin ollen siis ole verrattavissa keskenään. (Housen–Kuiken–Vedder 2012, 4.) Kompleksisuus esiintyy toisen kielen oppimisen tutkimuskirjallisuudessa kahdessa eri merkityksessä: kielellinen kompleksisuus ja kognitiivinen kompleksisuus. Kognitiivinen kompleksisuus on subjektiivista ja suhteellista, ja sillä viitataan toisen kielen ainesten prosessoinnin suhteelliseen vaikeuteen kieltä oppiessa, ja siihen vaikuttaa osin myös oppijan tausta. Kielellinen kompleksisuus on tärkeä osatekijä kognitiivisessa kompleksisuudessa, mutta ne eivät vastaa toisiaan. Kielellinen kompleksisuus on oppijasta riippumaton, ja se viittaa toisen kielen ainesten muodollisiin ja semanttisfunktionaalisiin ominaisuuksiin tai toisen kielen järjestelmien ominaisuuksiin. (mt., 4.)

Kompleksisuutta on tutkittu toisen kielen oppimisen tutkimuksessa monesta näkökulmasta, jotka voidaan karkeasti jakaa kahteen lähestymistapaan. Ensimmäinen niistä käsittää kompleksisuuden itsenäisenä muuttujana, jonka vaikutusta tutkitaan johonkin näkökulmaan toisen kielen taidossa tai suorituksessa. Esimerkiksi monissa tutkimuksissa on tutkittu jonkin kohderakenteen kompleksisuuden vaikutusta sen opettamiseen tai eri ohjeistusten tehokkuuteen. Toinen lähestymistapa käsittää kompleksisuuden riippuvana muuttujana ja usein sujuvuuden ja tarkkuuden kanssa. Tällöin se käsitetään toisen kielen suorituksen tai taidon yksinkertaisena kuvaajana. Toisen kielen oppijan suorituksen tai taidon kompleksisuutta mitataan kuvaamaan jonkin toisen muuttujan vaikutusta. Tällainen muuttuja voi olla esimerkiksi ikä, eri oppimiskonteksti tai erilainen ohjeistus. (Bulté–Housen 2012, 21–22.)

Yksi tapa mitata sanojen morfofonologista kompleksisuutta on laskea grafeemimuutosten määrä sanan vartalossa, kun sana muutetaan yksikön nominatiivista haluttuun muotoon, joka voisi olla esimerkiksi yksikön genetiivi. Tällöin jatkumon toisessa päässä voisi olla esimerkiksi sellainen sana kuin kauppa ja toisessa päässä esimerkiksi sana katteeton. Jälkimmäisessä sanassa tapahtuu sekä vokaali- että konsonanttivaihteluita. Tällöin jokainen grafeemin lisääminen, poistaminen tai vaihtaminen laskettaisiin yhdeksi muutokseksi. (Martin 1995, 89.)

Tässä tutkielmassa tarkastelen sanojen kompleksisuutta monien eri tekijöiden kannalta. Nämä tekijät liittyvät suurimmaksi osaksi sanojen rakenteeseen, mutta otan huomioon myös sanojen syntaktisen funktion. Sanoista tarkastellaan astevaihteluparia, sijamuotoa, sanan pituutta tavuina, sanatyyppejä ja syntaktista funktiota. Astevaihteluparin valinta, sijamuoto, tavujen määrä ja rakenne sekä sanatyyppi vaikuttavat kaikki sanan taivutukseen ja näin ollen myös astevaihteluun. Syntaktinen funktio itsessään ei vaikuta sanojen taivutukseen tai astevaihtelun käyttöön, mutta tietyn syntaktisen funktion mahdollinen sijamuoto vaikuttaa, jonka takia tätäkin on mielenkiintoista tarkastella. Kirjoitan näistä tarkemmin luvussa 3.1.

### 2.3 Astevaihtelu

Astevaihtelu on morfofonologista vaihtelua, joka koskee klusiileja *p*, *t* ja *k*, jos niitä edeltää soinnillinen äänne. Koska tutkielmassa tarkastellaan sanoja, joissa on *k*:n astevaihtelua, kaikki seuraavat esimerkit ovat sellaisia, joissa klusiili *k* vaihtelee. Astevaihtelu voi olla joko kvantitatiivista tai kvalitatiivista eli sitä on kahdenlaista. Kvantitatiivisessa vaihtelussa varioidaan klusiilin pituus (*kk* : *k*) ja kvalitatiivisessa vaihtelussa klusiilin laatu (esim. *k* : *v*). Sen lisäksi, että astevaihtelua on kahdenlaista, sillä on myös kaksi toteumaa, jotka ovat *heikko aste* ja *vahva aste*. (VISK § 41.) Kun sanassa on vahva aste, siinä esiintyy kvantitatiivisessa vaihtelussa geminaatta-*k* tai kvalitatiivisessa vaihtelussa yksittäinen *k*. Kun sana taas on heikossa asteessa, siinä on kvantitatiivisessa vaihtelussa yksittäinen *k* ja kvalitatiivisessa vaihtelussa esimerkiksi *v* tai jokin muu sen vaihtelupareista (vaihtelupareista lisää luvussa 3.1.1). Astevaihtelun yleisperiaatteena pidetään sitä, että yleensä avotavun edellä on vahva aste (*kak.ku*) ja umpitavun edellä heikko aste (*ka.kun*), mutta tästä on myös runsaasti poikkeuksia. (VISK § 43). Avotavu tarkoittaa sellaista tavua, jonka viimeinen äänne on vokaali, ja umpitavu puolestaan tarkoittaa sellaista tavua, jonka viimeinen äänne on konsonantti. Astevaihtelu voi olla joko suoraa tai käännteistä. Suorassa astevaihtelussa sanan perusmuodossa on vahva aste (*kakku*, *kyky*) ja taivutetussa muodossa heikko aste (*kakun*,

*kyyn*). Käänteisessä astevaihtelussa muodot ovat päinvastoin eli perusmuodossa on heikko aste (*kookus, palje*) ja taivutetussa muodossa vahva aste (*kookukaan, paljeen*). (VISK § 42.) Tässä tutkielmassa tutkin ainoastaan suoran astevaihtelun alaisia sanoja, mutta tarkastelen sekä kvantitatiivista että kvalitatiivista vaihtelua.

Taajuussanastossa on eniten *t:n* ja vähiten *p:n* astevaihtelun alaisia sanoja, eli *k:n* astevaihtelun alaisia sanoja on suhteellisen paljon, koska astevaihtelua ylipäänsä on noin kolmasosassa tuhannesta sanasta (VISK § 41). *Isossa suomen kieliopissa* on listattuna eri astevaihtelutyypin lekseemimäärät Perussanakirjassa. Tästä nähdään, että suoran astevaihtelun lekseemejä, joissa on *kk : k* -vaihtelua on sanakirjassa noin 1220 kappaletta, *k : j* -vaihtelua on 19 lekseemissä, *k : ø* -vaihtelua on 292 lekseemissä ja *k : v* -vaihtelua on viidessä sanassa. (VISK § 42). Näin ollen kvantitatiivisen vaihtelun alaisia sanoja on Perussanakirjassa siis 1220 ja kvalitatiivisen vaihtelun sanoja 316.

Äidinkielen puhuja ei yleensä tunnista vartaloiden äännevaihteluita, ellei hän ole kielen ammattilainen. Tämä voi olla siksi, että vartalon muutokset ovat äidinkieliselle puhujalle niin automaattisia tai siksi, että vartalon äänne muutoksilla ei ole mitään kieliopillista tehtävää. Tästä on esimerkkinä se, että sana, joka taipuu ilman vartalomuutoksia (*kissa : kissan : kissaa*) ilmaisee täysin samaa asiaa, kuin sana, jonka vartalossa on äänne muutoksia (*katti : katin : kattia*). Kielellisenä universaalina pidetään sitä, että kielellisillä muodoilla on jokin kieliopillinen tehtävä, mitä suomen kielen vartalomuutoksilla ei ole. Vartalomuutokset ovat kuitenkin säilyneet erilaisista historiallisista, kommunikatiivista ja systeemisistä syistä. (Martin 1995, 8.) Astevaihtelulla ei siis ole mitään kieliopillista tehtävää, mutta se on niin yleinen ilmiö, että sen opettaminen on tärkeää. Tätä voi perustella sillä, että astevaihtelun kohtaa monta kertaa melkein jokaisessa lauseessa (Karlsson 1982, 100). Tämä ei kuitenkaan tarkoita sitä, etteikö astevaihtelulla tai sen puuttumisella olisi mitään merkitystä, koska kuten edelläkin mainittu, astevaihtelu synnyttää minimipareja. Tämän lisäksi myös puhutussa kielessä, jossa sanoja usein lyhennetään eli päätteitä voi jäädä pois, astevaihtelu voi muuttaa sanan merkitystä.

## 2.4 Astevaihtelu suomi toisena kielenä -oppijoilla

Suomen kieli on morfologisesti runsas, ja morfologisilla ilmiöillä on merkittävä semanttinen tehtävä, eikä suomea sen vuoksi voi puhua ymmärrettävästi ilman, että hallitsee sen taivutussysteemiä. Sanoilla voi olla useitakin taivutusvartaloita, minkä takia morfologian oppiminen voi tuntua oppijasta hankalalta. (Martin 1990, 89–90.) Astevaihtelua on tutkittu

jonkin verran suomi toisena kielenä -kontekstissa esimerkiksi muutamissa pro gradu -tutkielmissa (ks. Lehtinen 1998; Hazāne 2009). Esimerkiksi Hazāne (2009) tutki latvialaisten opiskelijoiden nominitaivutusta ja hän havaitsi, että astevaihtelu on yksi virheisiin johtavista syistä.

Astevaihtelua on tutkittu osana suomenoppijoiden nominitaivutusta (Martin 1995). Suomenoppijoiden nominitaivutuksessa astevaihteluvirheitä tarkasteltiin siitä näkökulmasta, että astevaihtelusääntöjä käytettiin väärin tai ei käytetty ollenkaan. Tämä ei kuitenkaan tarkoita, että asia olisi aivan näin yksiselitteinen, mutta virheitä ja tuottamisen ongelmia oli helpompi tarkastella sääntöjen käyttämisen ja käyttämättömyyden avulla. Suomenoppijalle astevaihtelu ei ole pelkästään jokin tietty taivutusmuotoa tai avo- ja umpitavuja koskeva asia, vaan ennemminkin sanojen kirjoittamiseen ja lausumiseen vaikuttava asia. Tästä syystä virheitä tarkasteltiin erilaisten taivutusprosessien tuloksena. (Martin 1995, 153–154.) Nominitaivutukseen vaikuttavaksi tekijäksi nimettiin muun muassa sanan morfologinen kompleksisuus, vaihteluiden erottuvuus sekä sanan, muodon ja sanatyypin frekvenssi (Martin 1995, 268). Tulokseksi saatiin esimerkiksi se, että sanat voitiin asettaa vaikeusjärjestykseen niiden morfofonologisen kompleksisuuden perusteella (Martin 1995, 266).

Astevaihtelun esitystavat vaihtelevat tutkijakohtaisesti ja kontekstista riippuen. Esimerkiksi suomen opetuksen alkeissa laajalti käytössä oleva oppikirja *Suomen mestari 1* esittelee astevaihtelun ensimmäisen kerran jo toisessa kappaleessa, ja ilmiöstä käytetään nimitystä *k-p-t*-vaihtelu. Kirjassa opetetaan ensin suora astevaihtelu nominien osalta ja seuraavassa kappaleessa verbien osalta. Käänteinen astevaihtelu opetetaan vasta kappaleessa yhdeksän. Kirjassa astevaihtelun kuvauksessa nojataan vahvasti vaihteluparien esittelyyn sekä siihen, missä sijamuodoissa tai verbimuodoissa ja -tyypeissä sitä esiintyy. Poikkeukset esitellään myös lyhyesti. Suoraa ja käänteistä vaihtelua ei siis juuri eksplikoida, vaan vaihtelun suuntaa kuvataan vain nuolilla. (Gehring–Heinzmann 2020.) Toisessa suomen kielen oppikirjassa *Ahaa! 1*, astevaihtelua käsitellään ensimmäisen kerran vasta sivulla 240 toiseksi viimeisessä kappaleessa ennen kertaavaa kappaletta. Kirjassa käsitelystä ilmiöstä käytetään opetuksessa termiä astevaihtelu. Astevaihtelu esitellään heti niin, että se koskee sekä verbejä että nomineja. Myös suora ja käänteinen vaihtelu esitellään heti, mutta kirjassa näihin viitataan kirjaimilla A ja B, ja ne opetetaan kunnolla vasta myöhemmin. Myös tässä oppikirjassa nojataan siihen, missä sijoissa ja verbimuodoissa on kutakin vaihtelua. (Tikkanen–Tähtinen 2017.) Näin ollen kummassakaan suhteellisen uudessa oppikirjassa astevaihtelun opetuksessa ei nojata aiemmin esiteltyyn astevaihtelun yleisperiaatteeseen avo- ja umpitavuista.

Leila White (2019, 14–19, 146–149) käsittelee teoksessaan *Suomen kielioppia ulkomaalaisille* nominien ja verbien astevaihtelua. Astevaihtelusta käytetään termiä nominien konsonanttivaihtelu eli astevaihtelu (*k, p, t*-vaihtelu). Astevaihtelua kuvaillaan tärkeäksi sanan vartalon muutokseksi. Sen kuvaillaan tarkoittavan sitä, että sanan vartalossa voi tapahtua viimeisen tavun rajalla muutoksia niin, että ennen päätettä olevat konsonantit *k, p* tai *t* voivat muuttua tai kadota, kun sanan perään lisätään tietynlaisia päätteitä. Kirjassa kerrotaan myös, että sanoilla on vahva ja heikko vartalo. Konsonanttivaihteluiden kuvataan voivan muuttua kolmella tavalla: kahdesta konsonantista tulee yksi (*kk : k*), yksi konsonantti menee pois tai muuttuu (*k : ø*) tai *k, p* tai *t* muuttuu samanlaiseksi kuin edellinen konsonantti (*nk : ng*). (mt. 14–15.)

Teoksessa suoraa astevaihtelua kutsutaan tyyppin A vaihteluksi ja käänteistä tyyppin B vaihteluksi. Useimpien sanojen nominatiivissa kerrotaan olevan vahva konsonantti ja vartalossa heikko (tyyppi A) ja joidenkin sanatyypin nominatiivissa heikko konsonantti ja vartalossa vahva (tyyppi B). (White 2019, 16.) Konsonanttipareiksi nimetään kummankin vaihtelun osalta *kk : k, pp : p, tt : t, k : ø, p : v, t : d, nk : ng, mp : mm, lt : ll, nt : nn* ja *rt : rr*. Tyyppiin A kerrotaan kuuluvan sanat, joiden nominatiivimuodon lopussa on yksi vokaali, joka ei kuitenkaan ole *e*. Vahva konsonantti on aina yksikön nominatiivissa, yksikön ja monikon partitiivissa, yksikön ja monikon illatiivissa, yksikön ja monikon essiivissä ja monikon genetiivissä. Teoksessa kerrotaan, että heikko konsonantti on muissa sijamuodoissa. Teoksessa tuodaan esiin myös, että heikko vartalo on aina silloin, kun sijapäätteenä on yksi konsonantti (*-t, n*) tai yhdistelmänä kaksi konsonanttia ja vokaali (esim. *-lle, -ksi*) ja aina muuten vartalo on vahva. (mt. 17.) Tyyppiin B kuuluvat sanat, jotka loppuvat perusmuodossa joko *e*:hen tai konsonanttiin. Heikon konsonantin kerrotaan olevan yksikön nominatiivissa ja partitiivissa ja vahvan konsonantin muissa sijamuodoissa. Tavallisimpien tyyppiin B kuuluvien sanojen kerrotaan loppuvan joko *-e, -as, -tOn* tai *-in*. Kerrotaan myös, että *k : ø* voi vaihdella myös käänteisesti. Tämän lisäksi molemmista tyypeistä kerrotaan muutama poikkeus. (mt. 18.)

## 2.5 Eurooppalainen viitekehys

Eurooppalainen viitekehys (EVK) on apuväline, jonka tarkoitus on auttaa kolmella eri tavalla. Sen on tarkoitus auttaa ammattilaisia Euroopassa luomaan esimerkiksi yhtenäisiä opetussuunnitelmia ja oppikirjoja, auttaa ylittämään kommunikaatioesteitä, jotka johtuvat erilaisista koulutusjärjestelmistä Euroopassa sekä auttaa määrittämään taitotasot, joiden

mukaan voidaan määritellä ja seurata oppijoiden kielitaidon tasoja. EVK on julkaistu vuonna 2001, ja sen määrittelemiä taitotasoa tulisi soveltaa esimerkiksi aikuisten maahanmuuttajien kielitarpeisiin. EVK:n kielitaitotasot ovat keinotekoisia ja ne ovat vastineita tarpeelle tehdä esimerkiksi eksplisiittisiä oppimistavoitteita. Jos ajatellaan, että maahanmuuttajien tulisi hallita asuinmaansa kieltä, on tarpeellista määritellä millä taitotasolla kieltä tulisi osata. Euroopan maat tekevät tämän määrittelyn usein juuri EVK:n avulla, jossa on kuusi nousevaa tasoa, jotka ovat A1, A2, B1, B2, C1 ja C2. Näissä on määritelty erikseen kielen ymmärtäminen ja tuottaminen sekä vuorovaikutus. (Council of Europe, EVK, kielen taitotasot ja profiilit, 2023, sähköinen lähde.)

Suomessa on laadittu kielitaidon tasojen kuvausasteikko, joka on sovellus tästä Euroopan neuvoston kehittämästä yhteisestä Eurooppalaisesta viitekehuksesta. Taitotasoa on jaettu edelleen vielä niin, että A1-tasosta on tasot A1.1, A1.2 ja A1.3, A2-tasosta on tasot A2.1 ja A2.2, B1-tasosta on B1.1 ja B1.2, B2-tasosta on B2.1 ja B2.2 ja tasoista C1 ja C2 on tarkemmat erittelyt vain tasolta C1.1. Taitotasoa A1 kuvaillaan niin, että oppijalta onnistuu suppea viestintä kaikkein tutuimmissa tilanteissa. Taitotasolla A2 oppijalla tulisi olla perustarpeet välittömään sosiaaliseen kanssakäymiseen ja lyhyeen kerrontaan. Taitotasolla B1 oleva oppija kykenee selviytymään arkielämässä kohdekielellä ja taitotasolla B2 hän pystyy selviytymään säännöllisessä kanssakäymisessä syntyperäisten kanssa. Taitotasolla C1 ja C2 oppija kykenee selviytymään monissa vaativissa kielenkäyttötilanteissa. (Kielitaidon tasojen kuvausasteikko 2012, sähköinen lähde.)

Tässä tutkielmassa Eurooppalainen viitekehys on keskeisessä roolissa, koska astevaihtelun kehitystä seurataan taitotasoin. Kielitaidon tasojen kuvausasteikossa ei luonnollisesti mainita suoraan astevaihtelua, mutta sanojen virheellisyys mainitaan. Esimerkiksi tasolla A1.3 mainitaan, että alkeellisessa puheessa esiintyy peruskielioppivirheitä ja alkeellisessa vapaassa kirjoituksessa monenlaisia virheitä. Myös tasolla A2.1 mainitaan, että oppija tekee toistuvia virheitä perusasioissa, esimerkiksi taivutuksessa, kirjoittaessaan. Tasolla A2.2 sanotaan, että virheitä tulee enää harvinaisemmissa muodoissa ja rakenteissa, ja yksinkertaiset sanat ja rakenteet ovat virheettömiä. (Kielitaidon tasojen kuvausasteikko 2012, sähköinen lähde.)

## 3 Aineisto ja metodit

### 3.1 ICLFI-korpus

#### 3.1.1 Yleistä

Korpuukset mahdollistavat kielentutkijoille tutkimuksen, jossa kieltä voi tutkia suuren sähköisen aineiston avulla. Suomen kielen osalta on sellaisia korpuksia, jotka sisältävät äidinkielisten puhujien kieltä sekä sellaisia korpuksia, joissa on oppijansuomea. (Jantunen–Pirkola 2015, 88.) Tässä tutkielmassa käytetään oppijansuomea sisältävää korpusta, jonka nimi on Kansainvälinen oppijansuomen korpus eli ICLFI-korpus. ICLFI on lyhenne sanoista International Corpus of Learner Finnish. Korpus on suomi vieraana kielenä -korpus, jos noudatetaan määritelmää siitä, että suomi vieraana kielenä tarkoittaa ulkomailla toteutettavaa suomen opetusta. Korpusta on koottu vuodesta 2007 Oulun yliopiston johtamassa hankkeessa, jonka nimi on *Korpustutkimus oppijankielen kielikohtaisista ja universaaleista ominaisuuksista*, ja hankkeessa oli mukana yhteensä viisi yliopistoa. Korpuksessa on suomen kielen opiskelijoiden tekstejä, jotka opiskelevat suomea joko pääaineena tai sivuaineena, ja mukana on myös lyhyemmille kursseille osallistuneiden oppijoiden tekstejä. Opiskelijat opiskelevat yli 20 ulkomaisessa yliopistossa ja aineiston keräämisessä on auttanut näiden yliopistojen opetushenkilökunta. (Jantunen–Pirkola 2015, 89.) Korpuksessa on sekä käsin kirjoitettuja tekstejä, jotka on siirretty sähköisen muotoon sekä tekstejä, jotka on kirjoitettu sähköisesti tekstinkäsittelyohjelmalla. (Jantunen–Pirkola 2015, 91).

Vuonna 2011 ICLFI-korpuksessa on miljoona sanetta ja 5400 tekstiä. Tehtävät, joiden pohjalta tekstit on kirjoitettu, ovat opettajien määrittelemiä. Kerätyt tekstit ovat kokonaisia, eivätkä osia teksteistä. Oppijoilla, joilta tekstit on kerätty, on 22 eri äidinkieltä. Taitotasoiltaan oppijat ovat sekä alkeistason että keskitason oppijoita ja mukana on myös edistyneitä oppijoita. (Jantunen 2011, 89.) Korpuksen tekstit on annotoitu morfologisesti, ja ne on myös arvioitu eurooppalaisen viitekehysten taitotasoasteikon mukaisesti (ICLFI-korpuksen kuvailutiedot). Tekstit ovat esimerkiksi mielipidekirjoituksia, esseitä, päiväkirjoja, sähköpostiviestejä ja kirjeitä. Korpuksen tekstit ovat teemaltaan yleisiä, eli ne eivät kuulu spesifisti mihinkään erikoisalaan. (Jantunen–Pirkola 2015, 92). Ajallisesti korpus on synkroninen, mutta osalta oppijoista mukana on tekstejä useamman vuoden ajalta, jolloin korpus mahdollistaa osittain myös diakronisen tutkimuksen. Korpus sisältää tekstejä, joiden kirjoittajien kielitaito on hankittu oppimalla, mutta tämä ei kuitenkaan ole niin yksiselitteistä,

koska osa kirjoittajista on voinut myös omaksua kieltä esimerkiksi oleskelemalla Suomessa tai jos joku heidän lähipiiristään on puhunut heille suomea. (Jantunen 2011, 91).

Korpuksia pystyy hyödyntämään tutkimuksessa aina paremmin sen mukaan, kuinka paljon tekstien kirjoittajilta on kerätty taustatietoja ja miten tarkasti taustatiedot on dokumentoitu. ICLFI-korpukseseen on kerätty runsaasti taustamuuttujia oppijoista, oppimiskontekstista, tekstistä ja muista ominaisuuksista, ja niitä on yhteensä 22. (Jantunen 2011, 93).

Taustamuuttujia, joita ICLFI-korpukseseen on kerätty, ovat esimerkiksi oppijan ikä ja sukupuoli, kielitaidosta äidinkieli, muut oppijan hallitsemat kielet ja suomen opiskelun vuosimäärä, oppimiskonteksti (onko kotona puhuttu suomea) ja teksteistä EVK:n taitotaso, tekstin tekstilaji ja tehtävänanto sekä keräysaika. ICLFI-korpuksessa taustamuuttujat ovat listattuna kunkin tekstin yhteyteen. (Jantunen–Pirkola 2015, 94–95.)

Olen poiminut tutkielmani aineiston valmiista korpuksesta, jolloin minun ei henkilökohtaisesti ole tarvinnut hankkia tutkittavilta tutkimuslupia. Aineistossa on mukana paljon taustamuuttujia, mutta kenenkään tutkittavan nimeä ei ole saatavilla. En kuitenkaan hyödynnä tutkielmassani kirjoittajien taustoja, joten minun ei ole tarvinnut itse huolehtia tutkittavien anonymiteetistä. Korpukseseen pääsy on kuitenkin luvanvaraista, joten kuka tahansa ei pääse käsiksi aineistoon, jolloin minulla on kuitenkin vastuu siitä, että vain minä katselen aineistoa ja sieltä löytyviä tietoja. Korpus on luvanvarainen varmasti osin siksi, että se sisältää niin yksityiskohtaista tietoa tutkittavista. Keräämäni aineisto on taulukkolaskelmaohjelmassa, johon on poimittu vain sanat, tutkittavat muuttujat sekä konteksti. Näin ollen minun ei myöskään tarvitse pitää huolta siitä, että hävitän aineiston tarpeen vaatimalla tavalla tutkielman valmistuttua.

### 3.1.2 Aineiston poiminta korpuksesta

Olen poiminut tarkasteltavan aineiston ICLFI-korpuksesta Kielipankin Korp-hakuliittymän avulla. Käytin Korpissa laajennettua hakua, ja hakukriteereiksi valitsin neljä eri ehtoa. Kaikissa hauissa oli kaksi samaa ehtoa, jotka olivat se, että sana on yksikössä sekä se, että sana on nomini. Nominihakuehdolla aineistosta tuli muihinkin sanaluokkiin kuuluvia sanoja kuin substantiiveja, mutta otin mukaan tarkasteluun vain tähän sanaluokkaan kuuluvia sanoja. Tämän lisäksi muodostin hakutermin, jolla löydettiin sanoja, joissa voi olla *k*:n astevaihtelua. Tämä oli kvalitatiivisessa vaihtelussa  $[\backslash w]^*[aeiouy\ddot{o}n\ddot{l}r]k[aiouy\ddot{o}]$  ja kvantitatiivisessa vaihtelussa  $[\backslash w]^*[aeiouy\ddot{o}n\ddot{l}r]kk[aiouy\ddot{o}]$ . Hakutermeistä jätin pois tarkoituksella *hk*-yhdistelmän, koska siinä esiintyy niin marginaalisesti astevaihtelua, ettei sen sisällyttäminen

ollut järkevää. Neljäntenä hakuehtona oli taitotaso CEFR:n (EVK) mukaan. Tällöin hakuja tuli siis yhteensä 12. Kuten luvussa 3.1.6 mainitsin, otin mukaan sanoja vain lauseiksi tulkittavista yhteyksistä, koska yksi tarkasteltava muuttuja on syntaktinen funktio. Jätin pois myös puhuttelulisäykset samasta syystä. Näillä kriteereillä mukaan otettavat sanat sijoitin taulukkolaskelmaohjelmaan ja analysoin ne. Poimin sanat korpuksesta yksi kerrallaan, ja analysoin sanat manuaalisesti tarkastellen jokaisen sanan erikseen.

## 3.2 Aineiston sanoista tarkasteltavat tekijät

### 3.2.1 Määrällinen ja laadullinen vaihtelu sekä vaihteluparit

Astevaihtelupareja ja vaihtelun tyyppejä on ryhmitelty monin tavoin (vrt. esim. Karlsson 2009), mutta aion tässä tutkielmassa nojata *Ison suomen kieliopin* ryhmittelyyn. *Isossa suomen kieliopissa* (VISK § 41) *k:n* astevaihtelupareiksi on kvantitatiivisen vaihtelun osalta listattu *kk : k*, *ηkk : ηk*, *lkk : lk* ja *rkk : rk* (marginaalisesti esiintyy myös *gg : g*-vaihtelua, mutta se on jätetty tarkastelun ulkopuolelle tässä tutkielmassa) ja kvalitatiivisen vaihtelun osalta *ηk : ηη*, *lk : l ~ lj*, *rk : r ~ rj*, *hk : h ~ hj* ja *k : ∅ (~v)*. Käytännössä ja tutkielman kannalta vaihteluparit voidaan yksinkertaistaa niin, että kvantitatiivisessa vaihtelussa vaihtelupareina ovat *kk : k* ja kvalitatiivisessa vaihtelussa *k : j*, *k : v*, *k : ∅*, *k : ∅(j)* ja *nk : ng* (*ηk : ηη*). Yksinkertaistus on tehty *Ison suomen kieliopin* astevaihtelupareista niin, että *kk : k*-vaihtelupareiksi on laskettu kaikki esiteltyt kvantitatiivisen vaihtelun parit. Kvalitatiivisen vaihtelun osalta yksinkertaistus on tehty niin, että *k : j*-vaihteluksi lasketaan parit *lk : lj*, *rk : rj* ja *hk : hj*, *k : v*-vaihteluksi vain *k : v*-pari, *k : ∅*-vaihteluksi on laskettu parit *lk : l*, *rk : r*, *hk : h* (tämä tosin jätettiin tarkastelun ulkopuolelle sen marginaalisuuden takia, ks. VISK § 44) sekä *k : ∅*. Tämän lisäksi erikseen ovat *k : ∅(j)* ja *nk : ng*. Kaksi viimeistä paria olen ottanut tarkasteluun siitä yksinkertaisesta syystä, että tällaisia vaihtelupareja sisältäviä sanoja oli paljon, ja niiden tarkasteleminen on mielenkiintoista. Se on mielenkiintoista, koska esimerkiksi sanan *aika* taivutus menee niin, että sanaan tulee kirjain *j*, vaikka kyseessä ei ole *k : j*-vaihtelu. Sanan kirjain *i* muuttuu *j*:ksi, koska sen ympärillä olevat äänne saavat sen joka tapauksessa kuulostamaan siltä, kuin siinä olisi *j*, jos sen kirjoittaisi vain säännön mukaan muotoon *aian*. *Nk : ng*-vaihtelu on otettu erikseen mukaan, koska siinäkään ei tapahdu vain *k:n* katoamista, vaan äänne muuttuu. Yksinkertaistus voidaan tehdä, koska tutkimuksen fokuksessa ei ole tarkastella klusiilia edeltävää äännettä ja sen vaikutusta, vaan vain astevaihtelun osaamista klusiilin osalta ja muita tekijöitä. Käytännössä katsoen *Ison suomen kieliopin* vaihteluparit ovat kvantitatiivisessa vaihtelussa kaikki *kk : k*-vaihteluita ja

kvalitatiivisessa vaihtelussa joko  $k : j$  -,  $k : v$  - tai  $k : \emptyset$  -vaihteluita, koska klusiili  $k$  vaihtelee näiden esiteltyjen äänteiden tai puutosten kanssa, eikä edeltävä äänne vaikuta siihen, koska ne ovat kaikki soinnillisia, jolloin astevaihtelu toteutuu.

### 3.2.2 Sijamuodot

Tässä tutkielmassa olen ottanut mukaan tarkasteluun kaikissa sijamuodoissa olevat sanat. Instruktiiv- tai komitatiivisijaisia sanoja ei kuitenkaan ole, koska ne esiintyvät vain monikossa, ja tarkastelussa on vain yksikkömuotoisia sanoja. Osa sijamuodoista ovat niin marginaalisia, että kaikkia ei välttämättä löydy aineistosta. Yleisimmät sijat suomen yleiskielessä ovat nominatiivi, genetiivi ja partitiivi, ja yhdessä ne muodostavat kaksi kolmasosaa kaikista sijojen tekstiesiintymistä. Tämä selittyy sillä, että tavallisessa tekstilauseessa on yleensä nominatiivisijainen subjekti ja sitä luonnehtiva predikatiivi, joka on yleensä genetiivi tai partitiivi. Nominatiivi, genetiivi ja partitiivi voivat myös esiintyä sellaisessa lauseessa, jossa subjektin tekeminen kohdistuu johonkin, jolloin lauseessa voi olla partitiivi- tai genetiivisubjekti. Loput sijamuodot siis muodostavat kolmanneksen kaikista tekstiesiintymien sijoista, ja esimerkiksi sisäpaikallissijoja on selvästi enemmän kuin ulkopaikallissijoja. (VISK § 1227.)

Tutkielman hakukriteerien puitteissa mahdollisissa sanatyypeissä astevaihtelua on yhteensä kahdeksassa sijamuodossa 11:sta. Sellaiset sanat, joita kuuluisi taivuttaa ilman astevaihtelua, ovat essiivissä, partitiivissa ja illatiivissa. Näin ollen suurimmassa osassa sijamuodoista siis kuuluisi olla astevaihtelua, mikäli niissä olevissa sanoissa on  $k$  tai  $kk$  sopivassa yhteydessä. En ottanut tarkasteluun mukaan nominatiivimuotoisia sanoja, koska tarkastelen sanojen taivutusta, eikä nominatiivimuotoisia sanoja ole taivutettu.

### 3.2.3 Sanan pituus tavuina

EU:n LE-PAROLE-hankkeen yhteydessä koostettu Parole-korpus koostuu pääosin 90-luvun kirjoitetuista teksteissä, ja siinä on mukana muun muassa kirjoja, sanomalehtiä ja aikakauslehtiä. Tästä korpuksesta laskettuna kirjoitetussa suomenkielisessä tekstissä sanojen keskimääräinen pituus on 8,5 merkkiä. (Heikkinen–Lehtinen–Lounela 2001, kappale 4–7, sähköinen lähde.) Sanojen pituudet voivat vaikuttaa niiden oppimiseen niin, että mitä pidempi sana, sen vaikeampia sitä on oppia. Se ei kuitenkaan ole aivan yksiselitteistä, koska esimerkiksi jokin pitkä sana, joka koostuu tutuista äänneistä tai osista, voi olla helpompi oppia kuin vieras lyhyempi sana. (Malin 2012, 11.)

Tässä tutkielmassa tarkastelen sanojen pituutta kuitenkin tavujen määrän, enkä grafeemien määrän avulla. Tavulla tarkoitetaan yksikköä, joka voi koostua yhdestä vokaalista tai useammasta foneemista. Sananvartaloit ja siten myös sanat voivat koostua yhdestä tai monesta tavusta. Yksikin vokaali voi siis muodostaa tavun, mutta suomen kielen yleisimmät tavutyypit sisältävät sekä vokaaleita että konsonantteja (esimerkiksi CV tai CVC). (VISK § 11). Tavujen tarkastelu liittyy myös siihen, että astevaihtelun yleisperiaatteen mukaan avotavun edellä sanoissa on astevaihtelussa vahva aste ja umpitavun edessä heikko aste.

### 3.2.4 Sanatyypit

Sanatyypijaottelu, jonka olen korpuksesta poimituille nomineille tehnyt, on peräisin Leila Whiten (2019, 20–35) kirjasta *Suomen kielioppia ulkomaalaisille*. White on jakanut nominit 17 sanatyypiin (2019, 22–34). Suomen kielen sanat on mahdollista jakaa eri sanatyyppeihin sen mukaan, miten ne taipuvat eli millainen sanavartalo pitää muodostaa, ennen kuin siihen voi lisätä päätteitä. Sanatyypityksen perustana on sanan perusmuodon lisäksi sanavartalo ja sanan partitiivimuoto yksikössä. Sanoilla voi olla kaksi erilaista vartaloa: heikko ja vahva vokaalivartalo tai vokaalivartalo ja konsonanttivartalo. (White 2019, 20–21.) Sanatyypit on nimetty oppikirjasta löytyvien ”laatikoiden” avulla, joissa on esitetty aina yksi esimerkkisana, joka taipuu tällä tavalla. Yksinkertaisimman jaon voi tehdä sen perusteella, loppuuko sana vokaaliin vai konsonanttiin. Vokaaliloppuisten sanojen sanatyyppejä on kuusi ja konsonanttiloppuisten 11.

Samaisen Whiten teoksen astevaihtelun esitys on esitelty luvussa 3.2. Tässä on maininta siitä, että tyyppiin A, eli suoran astevaihtelun sanoihin kuuluvat sellaiset sanat, jotka loppuvat yhteen vokaaliin, joka ei voi olla *e*. Tyyppiin B, eli käänteisen astevaihtelun sanoihin kuuluvat sanat, jotka loppuvat joko konsonanttiin tai *e*:hen. Vokaaliloppuiset sanatyypit ovat KATTO, PANKKI, LEHTI, KIELI, UUSI ja HUONE (White 2019, 22–26). Konsonanttiloppuiset sanatyypit puolestaan ovat NAINEN, PUHELIN, TYÖTÖN, RAKAS, KAUNIS, VASTAUS, TOTUUS, VÄSYNYT, OLUT, SÄVEL ja TYTÄR (White 2019, 27–34). Koska tässä tutkielmassa tarkastellaan vain suoran astevaihtelun sanoja, esittelen seuraavaksi tarkemmin vokaaliloppuisten sanojen, pois lukien *e*-loppuiset sanat, sanatyypit.

KATTO-tyypin sanat loppuvat vokaaliin *a*, *o*, *u*, *y*, *ä* tai *ö*, eikä niiden loppuvokaali muutu vartalossa. Partitiivi muodostetaan lisäämällä perusmuodon perään *a* tai *ä* (*kattoa*). Näissä sanoissa astevaihtelu on mahdollinen (*katon*). Monikko muodostaessa *a* ja *ä* voivat muuttua. KATTO-tyypin sanoja ovat esimerkiksi *poika*, *katu*, *laukku* ja *jalka*. (White 2019, 22.) I-

loppuisia sanatyyppejä on neljä. Ensimmäinen on PANKKI-tyyppi, johon kuuluvissa sanoissa perusmuodon loppuvokaali ei muutu vartalossa. Myös PANKKI-tyypin sanoissa partitiivi muodostetaan lisäämällä perusmuodon loppuun *a* tai *ä* (*pankkia*). Astevaihtelu on näissä sanoissa mahdollinen (*pankin*). Monikkoa muodostaessa näistä sanoista vartalon *i* muuttuu *e*:ksi (*pankeissa*) lukuun ottamatta monikon genetiiviä. PANKKI-sanatyyppejä edustavat esimerkiksi sanat *sakki*, *kaupunki*, *kuppi* ja *tomaatti*. (White 2019, 23.) LEHTI-sanatyypin sanoissa perusmuodon *i* muuttuu vartalossa *e*:ksi. Partitiivi muodostetaan niin, että valitaan vahva vokaalivartalo, joka loppuu *e*:hen (*lehte-ä*). Näissä sanoissa astevaihtelu on mahdollinen (*lehdet*). Monikkoa muodostettaessa vartalon *e* katoaa (*lehtiä*). Tähän sanatyyppeihin kuuluvat esimerkiksi sanat *tähti*, *kaikki*, *lahti* ja *suomi*. (White 2019, 23–24.)

KIELI-tyypin sanoissa perusmuodon *i* vaihtuu *e*:ksi vartalossa. Näiden sanojen partitiivissa on konsonanttivartalo, eikä siinä ole *e*:tä (*kiel-tä*). Näissä sanoissa ei ole astevaihtelua. Monikossa vartalossa oleva *e* katoaa ja vain monikon *i* jää (*kielissä*). KIELI-sanatyypin kuuluvat esimerkiksi sanat *puoli*, *sieni*, *nuori* ja *veitsi*. (White 2019, 24.) UUSI-sanatyypin sanoissa on erikoisia vartalon konsonantin muutoksia, koska *s* muuttuu *d*:ksi ja edelleen *t*:ksi. Partitiivi muodostetaan *t*-loppuisella konsonanttivartalolla ja partitiivin päätteellä *-ta* (*uutta*). Monikon nominatiivi muodostetaan lisäämällä *d*:n sisältävän vokaalivartalon perään *-t* (*uudet*), ja muissa monikkomuodoissa päätteeseen voi laittaa suoraan sanan perusmuotoon. UUSI-sanaryhmään kuuluvat esimerkiksi sanat *kuukausi*, *käsi*, *tosi* ja *vuosi*. (White 2019, 25.) Käytännössä siis tutkielman kannalta oleelliset sanatyypit ovat KATTO, PANKKI ja LEHTI, koska vain niitä voi esiintyä hakukriteerien puitteissa.

### 3.2.5 Vartalon muutokset

Sanavartaloissa voi suomen kielessä esiintyä morfofonologisia vaihteluita, mikä tarkoittaa sitä, että yhdellä morfeemiin kuuluvalla äänneellä voi olla erilaisia foneemiedustumia erilaisissa sanoissa. Tällaisia samoissa kohdissa esiintyviä erilaisia morfeemeja kutsutaan toistensa allomorfeiksi. Tällaista morfofonologista muuntelua voi esiintyä sekä suffikseissa että sanojen vartaloissa. Suffikseissa morfofonologinen vaihtelu voi olla esimerkiksi vokaaliharmoniasta johtuva etu- ja takavokaalin vaihtelu taivutustunnuksissa (vrt. *jalassa* – *kädessä*) tai illatiivin vaihtoehtoiset tunnuksiset (vrt. *Turkuun* – *suohon* – *nesteseen*). Vartalossa tapahtuvaa morfofonologista vaihtelua on esimerkiksi erilaisten sanojen taivutusmuodoissa, kuten *puhelin* : *puhelime-n* tai *tarra* : *tarro-issa*. Keskeisimmät vaihtelut ovat juuri astevaihtelu ja sen lisäksi edellä esitetty vokaalivaihtelu taivutustunnusten edellä,

kuten sanassa *tarra*. Muutoksia loppuvokaalisto on aiheuttaa erityisesti suffiksit, joiden alussa on *-i*. Tällaisia ovat esimerkiksi monikon tunnus *-i*, vertailumuotojen johtimista *-in* sekä *-inen*-johtin. Sanoihin lisättävät johtimet aiheuttavat vartalovokaaleihin vaihteluita epäjohdonmukaisemmin kuin taivutustunnusten lisääminen. (VISK § 40.) Tässä tutkielmassa keskitytäänkin juuri taivutustunnusten aiheuttamiin vaihteluihin.

Taivutustunnusten aiheuttamia vokaalivaihteluita voi tapahtua yksittäisissä vokaaleissa *a*, *e*, *i* ja *ä* sekä niiden lisäksi diftongeissa ja pitkissä vokaaleissa. Vokaalivaihtelut voivat olla esimerkiksi *A : O* tai *A : -*, *A : e* tai *i : e* ja *e : -* tai *VV : V*. Vaihteluihin vaikuttavat esimerkiksi sanojen tavumäärät sekä sanojen historia. (VISK § 45–51.) Vokaalivaihteluiden lisäksi sanoissa voi esiintyä esimerkiksi *t : s* -vaihtelua (*vesi : vetenä*), *s : ks* -vaihtelua (*tulos : tuloksen*), *m : n* -vaihtelua (*lumi : lunta*), *s : t : d : ks* -vaihtelua (*tyhmyys : tyhmyyteen : tyhmyyden : tyhmyksien*) (VISK § 66). Tässä tutkielmassa mahdollisia vartalovaihteluita on vain *i : e* -vaihtelu, jota on LEHTI-sanatyypissä.

### 3.2.6 Syntaktinen funktio

Yksi tarkasteltava seikka aineistossa on myös se, vaikuttaako sanan syntaktinen funktio sen taivuttamiseen vaadittavan astevaihtelun oikeellisuuteen. Olen tulkinut sanat joko subjekteiksi, objekteiksi, adverbiaaleiksi, predikatiiveiksi tai e-subjekteiksi. Koska syntaktinen funktio on yksi tarkasteltavista asioista, en ole ottanut korpuksesta tarkasteluun sellaisia sanoja, jotka eivät ole osana rakennetta, jonka voi tulkita lauseeksi. Olen voinut tulkita sanat myös tietyn lauseenjäsenen edustajaksi, vaikka niiden sija poikkeaisi siitä sijasta, jonka lauseenjäsenen norminmukaisesti edellyttäisi. Tämä johtuu osaltaan myös siitä, että olen määritellyt sanojen syntaktisen funktion sen mukaan, minkälaisen lausekkeen osana ne ovat. Tarkastellut muodot eivät siis aina ole olleet yksinään jossakin syntaktisessa funktiossa, vaan usein osana kyseistä lauseketta. Teen näin, koska tutkin sanojen astevaihtelun käyttöä, eikä syntaktisten funktioiden oikeellisuutta.

## 3.3 DEMfad-malli

Jyväskylän yliopiston ja Soveltavan kielentutkimuksen keskuksen tutkimushankkeeseen *Toisen kielen oppimisen polut (Topling)* kuuluvan tutkimuksen teoreettisena tavoitteena on ollut yhdistää kolme lähestymistapaa, joita on tavallisesti lähestytty erillään toisen kielen oppimisen tutkimuksessa. Tämän yhdistetyn kehyksen nimi on DEMfad-malli ja sitä on alettu kehittää jo aiemmassa *Cefling*-hankkeessa. (Jyväskylän yliopisto, Topling, sähköinen lähde.)

*Cefling*-hankkeen tavoitteena on ollut tutkia toisen ja vieraan kielen oppijan kielen kehitystä eurooppalaisen viitekehyksen taitotasolta toiselle siirryttäessä (Jyväskylän yliopisto, Cefling, 2012, sähköinen lähde). Kolme yhdistettyä lähestymistapaa ovat kielellinen, viestintätaidollinen ja suorituksen kehittymisen seuraaminen. Suorituksen kehittymisen seuraamisen kolme keskeistä käsitettä ovat sujuvuus, tarkkuus ja kompleksisuus. Hankkeen tutkimuksen pääpainona on ollut distribuution ja kompleksisuuden määrittäminen. (Jyväskylän yliopisto, Topling, sähköinen lähde.)

DEMfad-mallia on käytetty monessa oppijansuomea käsittelevässä tutkimuksessa. Tätä mallia on käytetty muun muassa monessa pro gradu -tutkielmassa. Näissä tutkielmissa metodia on käytetty esimerkiksi tutkittaessa nesessiivirakenteita (Huida 2019), objektin rakennetyyppejä (Jussilainen 2019) ja kieltokonstruktioita (Eränummi 2019). Seilonen (2013) on myös käyttänyt väitöskirjassaan soveltuvien osin DEMfad-mallia kvantitatiivisessa osuudessaan. Väitöskirjassaan hän tutki epäsuoraa henkilöön viittaamista oppijansuomessa.

DEMfad-mallissa D (=domain) viittaa siihen konstruktion, jota tutkitaan, E (=emergence) viittaa siihen taitotasoon, jolla konstruktion ensi kerran esiintyy oppijan kielessä ja M (=mastery) viittaa siihen taitotasoon, jolla konstruktion jo hallitaan. Mallin kirjaimet fad tulevat sanoista *frequency* (käyttötaajuus), *accuracy* (tarkkuus) ja *distribution* (distribuutio). (Martin 2022, 84.) Kielitaidon kehittymistä kuvataan usein CAF-triadin (*complexity, accuracy and fluency*) kolmen ulottuvuuden avulla, jotka ovat suomeksi kompleksisuus, tarkkuus ja sujuvuus. Myös DEMfad-malli lähestyy kielen oppimista kolmesta ulottuvuudesta, mutta ulottuvuudet ovat hieman erit. DEMfad-mallin avulla kielenpiirteiden kehitystä voidaan verrata eri taitotasoilla, eri tutkittavissa konstruktiossa ja kielissä. (Martin–Mustonen–Reiman–Seilonen 2010, 58–59.)

DEMfad-mallin *domain* viittaa siis siihen kielenpiirteeseen, jota ollaan tarkastelemassa. Tämän tutkielman kontekstissa *domain* on siis klusiili *k:n* astevaihtelusääntöjen osaaminen sanoja taivuttaessa. Domain voi olla myös jokin muu kielenpiirre, kuten esimerkiksi jokin kielen rakenne tai jokin sanaston osa, eikä tutkittavan konstruktion määritelmä täten ole mihinkään teoriaan sidoksissa (Martin–Mustonen–Reiman–Seilonen 2010, 59). *Emergence* on siis se taitotaso, jolloin tutkittavaa kielenpiirrettä esiintyy oppijan kielessä ensimmäistä kertaa. Jos tutkimuksen kohteena olisi esimerkiksi paikallissijat, ensiesiintymiseksi riittäisi jonkin substantiivin käyttö sellaisella suffiksilla, jonka voisi tulkita paikallissijan päätteeksi (mt., 59). Tässä tutkielmassa se tarkoittaa sitä taitotasoa, jolla oppija on ensimmäistä kertaa taivuttanut

jotakin sanaa niin, että siinä kuuluisi olla  $k$ :n astevaihtelua. *Mastery* voidaan määritellä siksi taitotasoksi, jolla oppija käyttää kielenpiirrettä suunnilleen kohdekielen mukaisella tavalla. Tarkkuuden voidaan määritellä olevan riittävä, kun 80 % käytetyistä muodoista on oikeita, mutta rajaa voidaan siirtää tarvittaessa. Jakaumalle ja distribuutiolle ei ole määritelty samanlaista rajaa. (mt., 60.) Tutkielman kannalta *mastery* saavutetaan siis sillä taitotasolla, jolla 80 % käytetyistä taivutusmuodoista ovat oikeita  $k$ :n astevaihtelusääntöjen kannalta.

Mallin kirjain  $f$  tarkoittaa siis frekvenssiä eli käyttötaajuutta. DEMfad-mallissa käyttötaajuus liittyy sujuvuuteen niin, että oppijan kirjoittaman sanamäärän ja siihen sisältyvän tutkittavan kohteen käyttömäärän katsotaan kertovan sujuvuudesta. Näin ollen suuremman käyttötaajuuden voidaan katsoa korreloivan paremman sujuvuuden kanssa. Käyttötaajuus ei kuitenkaan ole sujuvuuden ainoa mittari. (Seilonen 2013, 30.) Tutkittavan ilmiön käyttötaajuus lasketaan tuhatta sanaa kohden (Martin–Mustonen–Reiman–Seilonen 2010, 60). Tällä pyritään välttämään sitä, että tekstien pituus vaikuttaisi tuloksiin, koska tekstit voivat olla eri taitotasoilla hyvinkin eri pituisia (Seilonen 2013, 31.) Tässä tutkielmassa käyttötaajuus on laitettu käytäntöön siten, että korpukselta lasketaan kaikki astevaihtelusanat ja niiden suhteellinen osuus korpuksen kaikista substantiiveista, jotka eivät ole nominatiivissa. Tässä tarkastellaan myös sitä, miten astevaihtelusanojen osuus muuttuu eri taitotasojen välillä.

DEMfad-mallissa kirjain  $a$  tarkoittaa tarkkuutta. Tarkkuutta ei voi tarkastella ilman vertailukohdetta, johon pyritään. Vertailukohtana toimii muoto tai rakenne, jonka samanikäinen ja saman koulutustaustan omaava natiivipuhuja muodostaisi. Kuitenkin natiivipuhujakin voisi tehdä samanlaisia virheitä huolimattomuuttaan. Vertailukohtana ei siis aina toimi kohdekielen normi. (Martin–Mustonen–Reiman–Seilonen 2010, 60.) Ajatellaan, että vartalonmuutokset ovat aikuiselle suomea ensikielenään puhuvalle henkilölle automaattisia, eikä niitä juuri huomata paitsi silloin, kun esimerkiksi lapsi tekee virheen tai joku taivuttaa jonkin harvinaisen sanan kielikorvan vastaisesti (Martin 1995, 7). Näin ollen tässä tutkielmassa voisi ajatella, että vertailukohtana toimii juuri kielen normi, eli astevaihtelusääntö, jota tulisi osata noudattaa säännön mukaisesti. Tässä tutkielmassa tarkkuutta tarkastellaan oikeiden muotojen määrällä kaikilla taitotasoilla erikseen sekä kaikkien tarkasteltavien muuttujien kannalta.

DEMfad-mallin kirjain  $d$  tarkoittaa distribuutiota eli jakaumaa. Distribuutio sisältää sekä variaation että kompleksisuuden piirteitä. Mallin luonnissa ei haluttu käyttää sanaa kompleksisuus, koska sen mittarit eivät sovi hyvin suomen kieleen kaikilta osin. (Seilonen

2013, 32.) Nämä kompleksisuuden mittarit ovat sellaisia, joita yleisesti käytetään toisen kielen oppimisen tutkimuksessa. Distribuution täytyykin ajatella olevan jonkinasteinen kattotermi erilaisille ilmiöille, joita tarkastellaan oppijan kielen kehityksessä (Martin–Mustonen–Reiman–Seilonen 2010, 61.) Tässä tutkielmassa jakaumaa tarkastellaan tarkasteltavien sanojen eri muuttujien määrällä eri taitotasolla, eli esimerkiksi tarkastellaan sitä, montako illatiivisijaista sanaa taitotasolla B1 on.

## 4 Astevaihtelusanojen kompleksisuuden vaikutus niiden oppimiseen

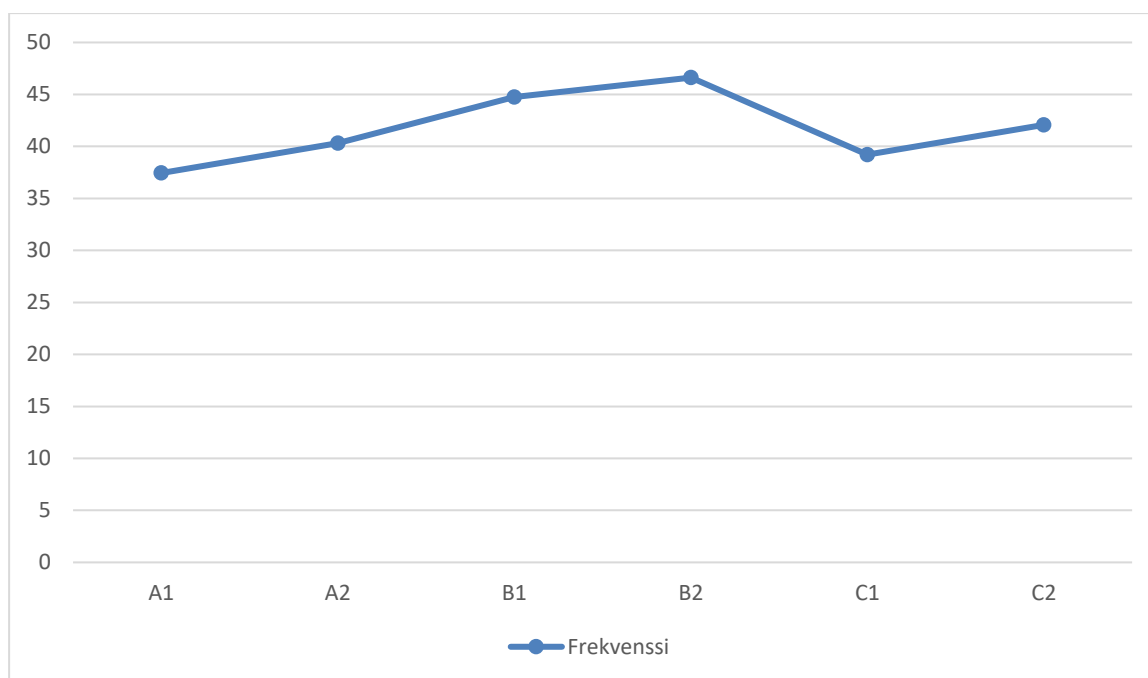
Tässä luvussa esittelen aineistoni analyysin pääosin määrällisesti, mutta teen huomioita myös laadullisesta näkökulmasta. Koska tutkimus on pääosin määrällinen, tarkastelen kullakin taitotasolla olevia identtisiä sanoja omina sanoinaan, vaikka se aiheuttaakin turhaa toistoa analyysissä. Taulukoissa esiintyvät prosenttiluvut on pyöristetty lähimpään tasalukuun pois lukien sellaiset luvut, jotka olivat tarkalleen kahden tasaluvun välissä, jolloin mukana on myös puolikkaita prosentteja. Huomionarvoista on, että pyöristämisen seurauksena joissakin taulukoissa muuttujien prosentuaalinen määrä koko otannasta ei ole yhteensä 100 %, vaan se voi olla joko hieman yli tai alle. Taulukkoteknisistä syistä sarakkeiden termillä NM tarkoitetaan norminmukaisuutta.

### 4.1 Kvalitatiivisen vaihtelun sanat

#### 4.1.1 Frekvenssi

DEMfad-metodin frekvenssi on tässä tutkielmassa kaikkien kvalitatiivisen astevaihtelun alaisten sanojen osuus koko kyseessä olevan taitotason substantiiveista, jotka ovat jossakin muussa sijamuodossa kuin nominatiivissa. Frekvenssi on laskettu tuhatta sanaa kohden, koska tarkastelen aineistoa taitotasoinen, ja korpuksessa on huomattavan erilaiset määrät sanoja jokaisella taitotasolla. Tällä yritetään minimoida aineiston määrän aiheuttamaa vääristymää. Ylimmille ja alimmille taitotasoinen on määritelty vähiten tekstejä, ja suurimmat määrät tekstejä on luokiteltu taitotasoinen B1 ja B2. Esimerkiksi taitotasoinen A1 on 1 095 substantiivina jossain muussa sijamuodoissa kuin nominatiivissa, kun taas taitotasoinen B1 niitä on 58 262.

Kuviosta 1 on nähtävissä frekvenssin kehitys kvalitatiivisen astevaihtelun sanojen osalta. Alkeistasoinen A1 frekvenssi on 37 ja A2-tasoinen se on 40. Keskitasoinen B1 ja B2 frekvenssit ovat 45 ja 47. Ylimmillä tasoinen C1 ja C2 frekvenssit ovat 39 ja 42. Frekvenssi nousee suhteellisen tasaisesti taitotasoinen A1 tasoinen B2, jonka jälkeen se laskee tasoinen C1. C2-tasoinen se kuitenkin nousee hieman, mutta on edelleen hieman alempana kuin tasoinen A2, joten frekvenssi ei kasva lineaarisesti alemmilta taitotasoinen ylemmille. Se ei kuitenkaan myöhemmillä taitotasoinen laske enää yhtä alas kuin alkeistasoinen A1.



Kuvio 1. Kvalitatiivisten astevaihtelusanojen frekvenssin kehitys tuhatta substantiivia kohden.

Tässä tapauksessa frekvenssi itsessään ei näin tarkasteltuna kerro oppijoiden astevaihtelun osaamisesta. Siitä se kuitenkin kertoo, että astevaihtelusanoja käytetään jokaisella taitotasolla, eli ilmiötä esiintyy ensimmäisen kerran jo aivan ensimmäisellä taitotasolla. Ilmiö itsessään siis on oppijoiden teksteissä käytössä jo ensimmäisellä taitotasolla, mutta jokainen mahdollinen variantti ei ole käytössä jokaisen tarkasteltavan tekijän kohdalla vielä ensimmäisellä taitotasolla. Frekvenssi on kvalitatiivisen vaihtelun osalta jokaisella taitotasolla suhteellisen samansuuruinen, koska ylimmän ja alimman frekvenssin välillä on vain noin kymmenen sanaa tuhatta sanaa kohden. Näin ollen oppijat tuskin ovat millään taitotasolla ainakaan tietoisesti vältelleet tai yrittäneet käyttää astevaihtelusanoja väkisin. Tosin astevaihtelu on niin yleinen ilmiö suomen kielessä, että astevaihtelusanojen vältteleminen on hankalaa, vaikka niin haluaisikin tehdä. Kuten olen aiemmin maininnut, suomen kielen taajuussanaston tuhannesta yleisimmistä sanoista vajaa kolmannes taipuu astevaihtelun alaisena.

Frekvenssin katsotaan kertovan sujuvuudesta niin, että mitä korkeampi frekvenssi on, sitä sujuvampaa oppijan kieli on. Tämä on kuitenkin monesti ongelmallista (ks. Esim. Kajander 2013, 175), eikä asia ole tässä tutkielmassakaan yksinkertainen. Astevaihtelusanojen käyttöön voivat vaikuttaa monet tekijät, eikä niiden käyttäminen tai käyttämättömyys välttämättä kerro sujuvuudesta. Toisaalta käänteisesti ajateltuna sanavaraston laajuus voisi mahdollistaa synonyymien käytön, jos oppija haluaisi vältellä juuri astevaihtelusanoja. Ajattelen, että tässä tapauksessa frekvenssi kertoo enemmänkin oppijoiden sanavarastosta ja siitä, kuinka paljon se

sisältää sellaisia sanoja, joissa klusiili  $k$  on sellaisessa asemassa, että siihen tulee kiinnittää huomiota taivutettaessa.

#### 4.1.2 Vaihteluparien jakauma ja tarkkuus

Vaihteluparit ovat hyvin keskeisessä osassa kvalitatiivista vaihtelua tarkasteltaessa, koska niiden osaamiseen koko vaihtelu perustuu. Totta kai huomiota tulee kiinnittää myös lisättävään päätteeseen, mutta se mitä muutetaan ja miksi se muutetaan, riippuu vaihtelupareista. Vaihteluparin valinta puolestaan riippuu sanan äänteellisestä rakenteesta ja sanatyypistä. Olen aiemmin esitellyt sen, mitkä vaihteluparit on laskettu mihinkin vaihteluun. Todettakoon kuitenkin vielä se, että käytännössä katsoen sekä  $k : \emptyset (j)$  - että  $nk : ng$  -parit ovat  $k : \emptyset$  -vaihteluita, mutta tässä tutkielmassa niitä tarkastellaan omina vaihtelupareinaan. Olen koonnut taulukkoon 1 kvalitatiivisen vaihtelun jakaumat ja tarkkuudet vaihteluparien osalta. Jakauma on siis se, montako prosenttia kullakin taitotasolla on tietyn vaihteluparin sanoja. Tarkkuus näkyy taulukossa tummennettuna kullakin taitotasolla norminmukaisista muodoista kunkin vaihteluparin osalta. Viimeisellä rivillä on kokonaistarkkuusprosentti jokaisella taitotasolla. Viimeisessä sarakkeessa on vielä kokonaisnorminmukaisuusprosentti jokaisesta vaihteluparista. Kaikista kvalitatiivisen vaihtelun alaisista sanoista oikein on siis 95 %. Kaikki tutkielman taulukot noudattavat samaa logiikkaa.

Taulukko 1. Kvalitatiivisen vaihtelun sanojen vaihteluparien jakauma ja tarkkuus.

	A1		A2		B1		B2		C1		C2		Yhteensä vaihteluparia	
	n=40		n=100		n=100		n=100		n=100		n=100			
	mää- rä	NM	mää- rä	NM	mää- rä	NM	mää- rä	NM	mää- rä	NM	mää- rä	NM	mää- rä	NM
<b>k : <math>\emptyset</math></b>	40 % (16)	<b>100</b> % (16)	37 % (37)	<b>86</b> % (32)	34 % (34)	<b>94</b> % (32)	41 % (41)	<b>100</b> % (41)	29 % (29)	<b>100</b> % (29)	35 % (35)	<b>100</b> % (35)	36 % (192)	<b>96</b> % (185)
<b>k : <math>\emptyset</math> (j)</b>	10 % (4)	<b>100</b> % (4)	12 % (12)	<b>100</b> % (12)	30 % (30)	<b>97</b> % (29)	35 % (35)	<b>100</b> % (35)	42 % (42)	<b>100</b> % (42)	41 % (41)	<b>100</b> % (41)	30 % (164)	<b>99</b> % (163)
<b>k : j</b>	-	-	-	-	2 % (2)	<b>50</b> % (1)	1 % (1)	<b>100</b> % (1)	-	-	3 % (3)	<b>100</b> % (3)	1 % (6)	<b>67</b> % (4)
<b>k : v</b>	-	-	-	-	-	-	2 % (2)	<b>100</b> % (2)	5 % (5)	<b>100</b> % (5)	3 % (3)	<b>100</b> % (3)	2 % (10)	<b>100</b> % (10)
<b>nk : ng</b>	50 % (20)	<b>65</b> % (13)	51 % (51)	<b>84</b> % (43)	34 % (34)	<b>94</b> % (32)	21 % (21)	<b>100</b> % (21)	24 % (24)	<b>96</b> % (23)	18 % (18)	<b>100</b> % (18)	31 % (168)	<b>89</b> % (150)
<b>Yh- teen- sä</b>		<b>82,5</b> % (33)		<b>87 %</b> (87)		<b>94 %</b> (94)		<b>100</b> % (100)		<b>99 %</b> (99)		<b>100</b> % (100)	100 % (540)	<b>95</b> % (512)

Taitotasolla A1 vaihteluparien jakauma on sellainen, että eniten eli 50 % on vaihteluparia  $nk : ng$ , toiseksi eniten on vaihteluparia  $k : \emptyset$  (40 %) ja vähiten eli 10 % on vaihteluparia  $k : \emptyset (j)$ . Tarkkuus taitotasolla A1 on kokonaisuudessaan 82,5 % ja se on alin tarkkuusprosentti koko kvalitatiivisen vaihtelun osalta. Vaihteluparien osalta tarkkuus on sellainen, että  $k : \emptyset$  - ja  $k : \emptyset (j)$  -vaihtelut ovat 100 % oikein ja  $nk : ng$  -vaihteluparista oikein on vain 65 %. Jos ajatellaan yleensä DEMfad-metodissa hallintaa osoittavaa 80 %:n rajaa, niin koko taitotasolla oppijat ovat jo hallinnan puolella, tosin eivät kovin vahvasti. Yksittäisistä vaihtelupareista hallinnan ulkopuolelle jää vain vaihtelupari  $nk : ng$ . Tällä taitotasolla A1 on siis 7 virheellistä muotoa. Ne ovat: *sänkyllä, sänkyllä, aurinkosta, kaupunissa, kaupunkilla, sänkyssä ja sänkyn*. Kuudessa sanassa on jätetty soveltamatta astevaihtelua ollenkaan, vaikka sitä olisi norminmukaisesti pitänyt käyttää, ja sanassa *kaupunissa* on tiputettu  $k$  pois, mutta tilalle ei ole laitettu norminmukaisesti kirjainta  $g$ . Johtopäätöksenä tästä voi tehdä sen, että taitotasolla A1 vaikein vaihtelupari on selvästi  $nk : ng$ , ja tyypillisin virhe on se, että astevaihtelua ei käytetä lainkaan.

Tasolla A2 jakauma on hyvin samanlainen kuin alemmakin taitotasolla. A2-tasolla 12 % prosenttia astevaihtelusanoista sisältää  $k : \emptyset (j)$  -vaihtelun, 37 % sisältää  $k : \emptyset$  -vaihtelun ja 51 % sisältää vaihteluparin  $nk : ng$ . Sen sijaan kokonaistarkkuus on tällä tasolla jo hieman korkeampi, eli 87 %. Yksittäisten vaihteluparien tarkkuusprosentit näyttävät kuitenkin erilaisilta kuin aiemmalla tasolla. Tällä tasolla vaihteluparin  $k : \emptyset (j)$  -sanoista on 100 % oikein,  $k : \emptyset$  -sanoista oikein on 86 % ja  $nk : ng$  -vaihtelun sanoista oikein on 84 %. Näin ollen verrattaessa 80 % -hallintatasoon tämän tason oppijat hallitsevat astevaihtelun käytettyjen vaihteluparien osalta kokonaan. Verrattaessa taitotasoa keskenään on kuitenkin todettava, että prosentuaalisiin määriin vaikuttaa väistämättä se, että taitotasolla A1 tarkastelussa on 40 sanaa, kun taas muilla tasoilla on kaikilla tarkastelussa sata sanaa.

Tällä tasolla A2 on yhteensä 13 virheellistä muotoa, jotka ovat  $k : \emptyset$ -vaihtelun osalta *ruokaksi, ruokaksi, alunsa, Turkussa* sekä *sikan*, ja  $nk : ng$  -vaihtelun osalta *kaupungiin, lenikin, aurinkossa, helsinkin, henkestä, kaupungiin, sänkyn ja sänkyllä*. Näissä sanoissa on kahdenlaisia virheitä: astevaihtelua ei ole käytetty silloin, kun olisi pitänyt ja astevaihtelua on käytetty silloin, kun sitä ei kuuluisi käyttää. Ensimmäiseen ryhmään kuuluvat sanat *ruokaksi, ruokaksi, Turkussa, sikan, lenikin, aurinkossa, helsinkin, henkestä, sänkyn ja sänkyllä*. Sanassa *lenikin* (leningin) tosin on muutenkin kirjoitusvirhe, eikä käytetty muoto ole

virheellinen pelkästään astevaihtelun näkökulmasta. Jälkimmäiseen ryhmään kuuluvat siis sanat *alunsa*, *kaupungiin* ja *kaupungiin*. Näin ollen tällä tasolla vaikein vaihtelu oppijoille on ollut  $nk : ng$  ja tyypillisin virhe on se, että astevaihtelua ei sovelleta silloin, kun sitä kuuluisi soveltaa.

Taitotasolla B1 jakauma on jo hieman erilainen kuin aiemmilla taitotasolla. Tällä tasolla esiintyy ensimmäisen kerran myös uusi vaihtelupari, jota ei ole käytetty aiemmilla tasoilla. Jakauma on tällä tasolla sellainen, että  $k : j$ -sanoja on 2 %,  $k : \emptyset (j)$ -sanoja on 30 %, ja sekä  $k : \emptyset$  - että  $nk : ng$ -sanoja on 34 %. Kokonaistarkkuusprosentti on tälläkin tasolla korkeampi kuin aiemmalla tasolla, koska se on jo 94 %. Vaihteluparien tarkkuusprosentit ovat: 97 % ( $k : \emptyset (j)$ ), 94 % ( $k : \emptyset$  ja  $nk : ng$ ) ja uuden vaihteluparin  $k : j$  osalta tarkkuus on 50 %. Tosin viimeiseksi mainittuja esiintymiä on vain kaksi, joista vain toinen on norminmukainen, eikä kahden sanan otanta välttämättä vielä kerro luotettavasti ilmiön hallinnasta. Kuitenkaan siis vaihtelupari  $k : j$  ei ole tällä tasolla vielä hallinnassa 80 %:n rajaan verrattuna.

Tasolla B1 on kuusi virheellistä sanamuotoa. Nämä sanat ovat  $k : \emptyset$ -vaihtelun osalta *sikan* ja *Kaurismäkin*,  $k : \emptyset (j)$ -vaihtelun osalta *poikan*,  $k : j$ -vaihtelun osalta *jälkestä* ja  $nk : ng$ -vaihtelun osalta *Helsingiin* ja *sängyä*. Kaikissa muissa muodoissa astevaihtelua ei ole käytetty, mutta kahdessa viimeiseksi mainitussa astevaihtelua on käytetty, vaikka norminmukaisesti näin ei olisi kuulunut tehdä. Prosentuaalisesti vaikein vaihtelu tällä tasolla on siis  $k : j$ -vaihtelu, mutta määrällisesti sekä  $k : \emptyset$  - että  $nk : ng$ -vaihtelun sanoissa oli yksi virhe enemmän. Tälläkin tasolla tyypillisin virhe on astevaihtelun käyttämättä jättäminen.

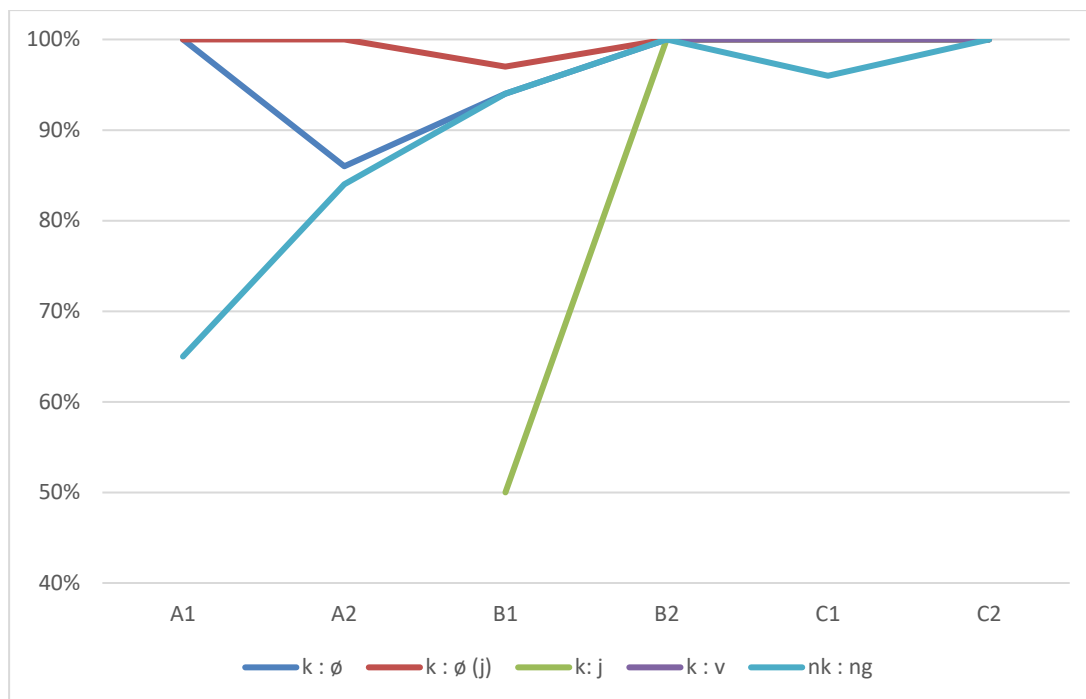
Seuraavalla tasolla B2 esiintyy ensimmäisen kerran viimeinen vaihtelupari, jota tutkielmassa tarkastellaan, joka on siis  $k : v$ -vaihtelu. Jakauma on tällä tasolla sellainen, että eniten eli 41 % sanoista on  $k : \emptyset$ -vaihtelua, toiseksi eniten eli 35 % on vaihtelua  $k : \emptyset (j)$ , kolmanneksi eniten eli 21 % on  $nk : ng$ -vaihtelua, toiseksi vähiten eli 2 % on uutta  $k : v$ -vaihtelua ja vähiten eli 1 % on  $k : j$ -vaihtelua. Kokonaisnorminmukaisuusprosentti on tällä tasolla ensimmäisen kerran 100 % eli tällä tasolla ei ole yhtäkään virheellistä muotoa. Tällä tasolla kaikki vaihteluparit ovat siis jo täysin hallinnassa, tosin 1 % ja 2 % osuus kaikista sanoista ei ole kovin luotettava indikaattori ilmiön hallinnasta.

Taitotasolla C1 esiintyy kaikkia muita vaihteluita paitsi kahdella aiemmalla tasolla jo esiintynyttä vaihtelua  $k : j$ . Jakauma on seuraavanlainen:  $k : \emptyset (j)$  sanoja on 42 %,  $k : \emptyset$ -sanoja on 29 %,  $nk : ng$ -sanoja on 24 % ja  $k : v$ -sanoja on 5 %. Norminmukaisuusprosentti on tällä tasolla vain hieman alhaisempi kuin edellisellä tasolla, ja se on 99 %. Muiden

vaihteluiden osalta norminmukaisuusprosentti on 100 %, mutta vaihtelun  $nk : ng$  -vaihtelun norminmukaisuus on 96 %. Tälläkin tasolla on kuitenkin vain yksi virheellinen muoto, joka on *Helsinki*. Tässä muodossa ei ole käytetty astevaihtelua, vaikka olisi pitänyt. 80 %:n rajaan verrattuna kuitenkin kaikki vaihteluparit ovat tällä tasolla jo hallinnassa.

Korkeimmalla tasolla C2 esiintyy kaikkia tarkastelussa olevia vaihtelupareja. Jakauma on tällä tasolla sellainen, että  $k : \emptyset (j)$  -sanoja on 41 %,  $k : \emptyset$  -sanoja on 35 %,  $nk : ng$  -sanoja on 18 % ja sekä  $k : j$  - että  $k : v$  -sanoja on 3 %. Korkeimmalla tasolla norminmukaisuusprosentti on taas täydet 100 %. Näin ollen tällä tasolla ei ole yhtäkään virheellistä muotoa ja kaikki muodot ovat 80 %:n rajan mukaan hallinnassa.

Vaihteluparien suurin jakauma on taitotasolla B2 ja C2, koska kummallakin tasolla on kaikkia tarkasteluja vaihtelupareja. Taulukosta 1 näkee myös vaihteluparikohtaiset kokonaisnorminmukaisuusprosentit. Ainoa vaihtelupari, joka on kokonaisuudessaan oikein on  $k : v$  -vaihtelu. Tämä on oppijoiden teksteissä kuitenkin vain kolmella viimeisellä taitotasolla, ja näitä sanoja on otannassa vain kymmenen. Muutenkin  $k : v$  -vaihtelu on siinä mielessä poikkeuksellinen, että sitä esiintyy vain viidessä eri sanassa suomen sanastossa (VISK § 41). Toiseksi norminmukaisimmin käytetty vaihtelu on  $k : \emptyset (j)$ , jonka norminmukaisuusprosentti on 99, ja siinä esiintyi vain yksi virheellinen muoto. Kolmanneksi eniten oikeita muotoja suhteessa määrään on  $k : \emptyset$  -vaihtelun sanoissa, ja oikeita muotoja on 96 %. Tämän jälkeen  $nk : ng$  -sanoja oikein on 89 %. Pienin norminmukaisuusprosentti on vaihteluparissa  $k : j$ , eikä sitäkään käytetä jokaisella taitotasolla. Näitä sanoja on otannassa vain kuusi, joten jo kaksi virheellistä muotoa laskee norminmukaisuusprosentin 67 %:iin. Kuviossa 2 on näkyvissä vaihteluparikohtaiset norminmukaisuudet taitotasoittain käyrinä. Kuten kuviosta näkyy, kaikki vaihteluparit ovat suhteellisen hyvin hallinnassa jo alusta alkaen. Hallinnan osoittaman 80 %:n rajan alittavat vain vaihteluparit  $nk : ng$  ja  $k : j$ , ja nekin vain ensimmäisellä taitotasolla, jolla niitä esiintyy oppijoiden teksteissä.



Kuvio 2. Kvalitatiivisen vaihtelun sanojen vaihteluparien norminmukaisuuden kehitys taitotasoin. Kuvio 2 näkee sen, miten eri vaihteluparit osataan eri taitotasoin.  $k : v$ -vaihtelu osataan jokaisella taitotasolla, jolla sitä on, täydellisesti. Vaihtelupari  $k : \emptyset(j)$  osataan myös suhteellisen tasaisesti jokaisella taitotasolla.  $k : \emptyset$ -vaihtelupari tekee heti toisella taitotasolla pienen notkahduksen, jonka jälkeen se nousee nopeasti taas 100 %:iin. Selkeimmät kehityskaaret on näkyvissä vaihteluissa  $k : j$  ja  $nk : ng$ . Näistä ensimmäinen nousee heti ensiesiintymistä seuraavalla tasolla 100 %:iin, mutta toinen nousee pikkuhiljaa pienen notkahduksen kanssa lopulta 100 %:iin taitotasolta toiselle edetessä.

#### 4.1.3 Sijamuotojen jakauma ja tarkkuus

Vaihteluparien lisäksi sijamuodot ovat hyvin keskeisessä osassa astevaihtelun käytössä. Astevaihtelua tapahtuu sanoja taivuttaessa, ja luvun lisäksi nominien taivuttaminen perustuu sijapäätteiden lisäämiseen. Tietyt sijapäätteet laukaisevat astevaihtelun, jonka takia sijamuotojen lähempi tarkastelu on mielenkiintoista astevaihtelun oppimista tutkittaessa. Otannassa on mukana myös possessiivisuffiksillisia ja liitepartikkeleja sisältäviä muotoja, joita koskee omat sääntönsä. Lähtökohtaisesti niitä liitetään kuitenkin sijapäätteisiin, mutta ne ovat silti tarkastelussa mukana. Taulukkoon 2 on koottuna kaikkien otannassa mukana olevien sijamuotojen ja possessiivisuffiksillisten ja liitepartikkeleja sisältävien muotojen tarkkuus ja jakauma.

Taulukko 2. Kvalitatiivisen vaihtelun sanojen sijamuotojen jakauma ja tarkkuus.

	A1		A2		B1		B2		C1		C2		Yhteensä sijamuotoa	
	n=40		n=100		n=100		n=100		n=100		n=100		määrä	NM
<b>Gene- tiiv</b>	Mää- rä 12,5 % (5)	NM <b>80 %</b> (4)	Mää- rä 25 % (25)	NM <b>84 %</b> (21)	Mää- rä 17 % (17)	NM <b>82 %</b> (14)	Mää- rä 30 % (30)	NM <b>100 %</b> (30)	Mää- rä 26 % (26)	NM <b>96 %</b> (25)	Mää- rä 35 % (35)	NM <b>100 %</b> (35)	26 % (138)	<b>93 %</b> (129)
<b>Parti- tiiv</b>	45 % (18)	<b>100 %</b> (18)	28 % (28)	<b>100 %</b> (28)	40 % (40)	<b>97,5 %</b> (39)	23 % (23)	<b>100 %</b> (23)	24 % (24)	<b>100 %</b> (24)	15 % (15)	<b>100 %</b> (15)	27 % (148)	<b>99 %</b> (147)
<b>Inessiivi</b>	17,5 % (7)	<b>71 %</b> (5)	24 % (24)	<b>92 %</b> (22)	18 % (18)	<b>100 %</b> (18)	19 % (19)	<b>100 %</b> (19)	22 % (22)	<b>100 %</b> (22)	11 % (11)	<b>100 %</b> (11)	19 % (101)	<b>96 %</b> (97)
<b>Elatiivi</b>	2,5 % (1)	<b>0 %</b>	4 % (4)	<b>75 %</b> (3)	6 % (6)	<b>83 %</b> (5)	4 % (4)	<b>100 %</b> (4)	11 % (11)	<b>100 %</b> (11)	7 % (7)	<b>100 %</b> (7)	6 % (33)	<b>90 %</b> (30)
<b>Illatiivi</b>	7,5 % (3)	<b>100 %</b> (3)	9 % (9)	<b>78 %</b> (7)	8 % (8)	<b>87,5 %</b> (7)	17 % (17)	<b>100 %</b> (17)	11 % (11)	<b>100 %</b> (11)	14 % (14)	<b>100 %</b> (14)	11 % (62)	<b>95 %</b> (59)
<b>Ades- siivi</b>	7,5 % (3)	<b>0 %</b> (0)	6 % (6)	<b>83 %</b> (5)	6 % (6)	<b>100 %</b> (6)	2 % (2)	<b>100 %</b> (2)	1 % (1)	<b>100 %</b> (1)	-	-	3 % (18)	<b>78 %</b> (14)
<b>Abla- tiiv</b>	2,5 % (1)	<b>100 %</b> (1)	-	-	1 % (1)	<b>100 %</b> (1)	-	-	-	-	-	-	0,4 % (2)	<b>100 %</b> (2)
<b>Trans- latiivi</b>	-	-	2 % (2)	<b>0 %</b> (0)	-	-	1 % (1)	<b>100 %</b> (1)	-	-	1 % (1)	<b>100 %</b> (1)	0,7 % (4)	<b>50 %</b> (2)
<b>Allatiivi</b>	-	-	-	-	2 % (2)	<b>100 %</b> (2)	-	-	1 % (1)	<b>100 %</b> (1)	7 % (7)	<b>100 %</b> (7)	2 % (10)	<b>100 %</b> (10)
<b>Essiivi</b>	-	-	-	-	-	-	2 % (2)	<b>100 %</b> (2)	2 % (2)	<b>100 %</b> (2)	3 % (3)	<b>100 %</b> (3)	1 % (7)	<b>100 %</b> (7)
<b>Genetii- vi + posses- siivi- suffiksi</b>	5 % (2)	<b>100 %</b> (2)	2 % (2)	<b>50 %</b> (1)	1 % (1)	<b>100 %</b> (1)	-	-	-	-	-	-	0,9 % (5)	<b>80 %</b> (4)
<b>Allatiivi + posses- siivi- suffiksi</b>	-	-	-	-	1 % (1)	<b>100 %</b> (1)	-	-	-	-	-	-	0,2 % (1)	<b>100 %</b> (1)
<b>Partiitii- vi + posses- siivi- suffiksi</b>	-	-	-	-	-	-	2 % (2)	<b>100 %</b> (2)	1 % (1)	<b>100 %</b> (1)	5 % (5)	<b>100 %</b> (5)	1 % (8)	<b>100 %</b> (8)
<b>Adessii- vi + posses- siivi- suffiksi</b>	-	-	-	-	-	-	-	-	1 % (1)	<b>100 %</b> (1)	-	-	0,2 % (1)	<b>100 %</b> (1)
<b>Elatiivi + posses- siivi- suffiksi</b>	-	-	-	-	-	-	-	-	-	-	1 % (1)	<b>100 %</b> (1)	0,2 % (1)	<b>100 %</b> (1)
<b>Partiitii- vi + partik- keli- -kaan</b>	-	-	-	-	-	-	-	-	-	-	1 % (1)	<b>100 %</b> (1)	0,2 % (1)	<b>100 %</b> (1)
<b>Yhteen- sä</b>		<b>82,5 %</b> (33)		<b>87 %</b> (87)		<b>94 %</b> (94)		<b>100 %</b> (100)		<b>99 %</b> (99)		<b>100 %</b> (100)		

Kuten taulukosta 2 näkyy, taitotason A1 jakauma on seuraavanlainen: partitiivimuotoisia sanoja on 45 %, inessiivimuotoisia sanoja on 17,5 %, genetiivimuotoisia 12,5 %, illatiivi- ja adessiivimuotoisia sanoja on 7,5 %, genetiivi + possessiivisuffiksi -muotoja on 5 % ja elatiivi- ja ablatiivimuotoisia sanoja on 2,5 %. Norminmukaisuus tällä tasolla on 82,5 %, joka koostuu siitä, että genetiivimuotoisista sanoista oikein on 80 %, inessiivimuotoisista 71 %, elatiivimuotoiset ja adessiivimuotoiset muodot ovat kaikki väärin ja loput tällä taitotasolla mukana olevista muodoista on oikein. Nämä kaikki edellä mainitut muodot esiintyvät siis ensimmäisen kerran ensimmäisellä taitotasolla A1, eikä esimerkiksi translatiivi-, allatiivi- tai essiivimuotoisia sanoja ole ollenkaan. Koko otannassa ei ole mukana yhtäkään abessiivisijaista sanaa, jolloin kyseistä sijamuotoa ei luonnollisesti ole taulukossakaan.

Virheellisiä muotoja on tällä tasolla siis edelleen seitsemän. Genetiivisijainen norminvastainen muoto on *sänkyn*. Muodossa on genetiivin päätte *-n*, jonka lisättäessä sanassa kuuluisi olla astevaihtelua. Tässä muodossa sitä ei kuitenkaan ole, vaan päätte on lisätty vain nominatiivimuodon perään tekemättä tarvittavia muutoksia. Inessiivisijaiset norminvastaiset muodot ovat *kaupunissa* ja *sänkyyssä*. Inessiivin päätte on *-ssA*, jolloin sanassa kuuluisi olla astevaihtelua. Käytetty muoto *kaupunissa* on väärin astevaihtelun kannalta, koska näyttää siltä, että *k* on jätetty pois, kuten *k : ø* -vaihtelussa kuuluukin tehdä, mutta kyseessä on *nk : ng* -vaihtelun sana, jolloin oikea astevaihtelullinen muoto olisi *kaupungissa*. Käytetyssä muodossa *sänkyyssä* päätte on lisätty perusmuotoon tekemättä tarvittavia muutoksia sanan vartalossa. Elatiivissa oleva väärä muoto on *aurinkosta*, ja adessiivisijaiset väärät muodot ovat *sänkylällä*, *sänkylällä* ja *kauppunkilla*. Elatiivin päätte on *-stA*, jolloin sanassa kuuluisi olla astevaihtelua, mutta päätte on lisätty sanan perusmuotoon, eikä astevaihtelua ole käytetty. Adessiivin päätte on *-lla*, jolloin näissäkin sanoissa kuuluisi olla astevaihtelullinen muoto, mutta *sänkylällä*-muodoissa päätte on lisätty perusmuotoon. *Kauppunkilla*-muoto on norminvastainen kahdella tavalla, koska siinä on sekä kirjoitusvirhe että astevaihteluvirhe, koska *nk*-yhdistelmää ei ole vaihdettu oikeaan muotoon *ng*.

Taitotasolla A1 vaikeimmat sijamuodot ovat sekä elatiivi että adessiivi, koska niiden kaikki muodot ovat norminvastaisia. Tosin näitä on yhteensä vain neljä, joten ihan näin suoria johtopäätöksiä ei luotettavasti voi tehdä. Määrällisesti yhtä vaikea sijamuoto on inessiivi, koska siinä muodossa olevista sanoista kaksi on väärin. Mielenkiintoista on se, että jo alimmalla tasolla on osattu muodostaa astevaihteluttomat muodot kaikista sen sijaisista sanoista, joissa ei kuuluisikaan olla astevaihtelua. Tähän kategoriaan kuuluvat siis partitiivi- ja illatiivisijaiset sanat sekä genetiivi + possessiivisuffiksi -muotoinen sana. Näissä sanoissa oli

kuitenkin kaksi sellaista muotoa, joissa oli muuta huomionarvoista: *väkensä* ja *sänkytä*.

*Väkensä*-muodossa ei ole tehty vartalovokaalin muutosta *i*:stä *e*:ksi, kuten sanassa kuuluisi tehdä, mutta astevaihtelun kannalta sana on oikeassa muodossa, koska *nsA*-suffiksin kanssa ei kuuluisikaan käyttää astevaihtelua. *Sänkytä* muoto on astevaihtelun kannalta oikein, koska sana on partitiivissa, mutta partitiivin päätte on väärä, koska tässä sanassa sen kuuluisi olla *-ä*.

Taitotasolla A2 jakauma on hieman erilainen kuin aiemmalla taitotasolla. Tällä tasolla on melkein sama määrä partitiivi- (28 %), genetiivi- (25 %) ja inessiivisijaisia (24 %) sanoja. Sen jälkeen eniten on illatiivisijaisia (9 %) sanoja, adessiivisijaisia (6 %) ja elatiivisijaisia (4 %) sanoja. Tällä taitotasolla on yhtä paljon (2 %) sekä translatiivisijaisia että genetiivi + possessiivisuffiksi -muotoja. Tällä tasolla esiintyy siis ensimmäisen kerran sana, joka on translatiivissa. Norminmukaisuusprosentti on tällä taitotasolla 87 %. Tälläkin tasolla kaikki partitiivisijaiset sanat ovat oikein, inessiivimuotoja on oikein 92 % ja genetiivimuotoja on oikein 84 %. Adessiiveista oikein on 83 %, illatiiveista aiemmasta poiketen 78 % ja elatiiveista jo enemmän eli 75 %. Tälläkin tasolla ensiesiintymisensä tehnyt sijamuoto eli tässä tapauksessa translatiivi on kokonaan norminvastainen. Tällä tasolla genetiivi + possessiivisuffiksi -muodoista on 50 % oikein. Tällä tasolla ei kuitenkaan ole jo aiemmin esiintynyttä ablatiivimuotoa.

Tällä tasolla vaikein sijamuoto on prosentuaalisesti translatiivi (0 %), mutta määrällisesti eniten virheellisiä muotoja on genetiivisijaisissa sanoissa (4). Tämän tason virheellisistä muodoista neljä on genetiivissä (*sikan*, *lenikin*, *helsinkin* ja *sänkyn*), kaksi inessiivissä (*Turkussa* ja *aurinkossa*), yksi elatiivissa (*henkestä*), kaksi illatiivissa (*kaupungiin* ja *kaupungiin*), yksi adessiivissa (*sänkyllä*), kaksi translatiivissa (*ruokaksi* ja *ruokaksi*) ja yksi on genetiivi + possessiivisuffiksi -muotoinen (*alunsa*). Genetiivimuotoiset virheelliset muodot ovat kaikki sellaisia, että astevaihtelua ei ole käytetty, vaikka *n*-päätteen lisääminen olisi sitä edellyttänyt, tosin muodossa *lenikin* on myös kirjoitusvirhe. Myös inessiivi-, elatiivi- ja adessiivisijaiset sanat ovat väärin samasta syystä. Translatiivissa on kaksi sanaa, jotka ovat samanmuotoiset, joten niissä on samanlaiset virheet, eli astevaihtelua ei ole käytetty, vaikka päätte *-ksi* sitä vaatisi. Tällä tasolla on virheellisiä muotoja sellaisissakin muodoissa, joissa ei kuuluisi olla astevaihtelua, ja nämä ovat kaksi samanmuotoista illatiivisijaista sanaa *kaupungiin* sekä genetiivi + possessiivisuffiksi -muoto *alunsa*. Molemmat virheelliset muodot ovat siinä mielessä ymmärrettäviä, että jos sanoja tarkastelee astevaihtelun pääsäännön valossa, näissä kuuluisikin käyttää astevaihtelua. Näitä muotoja muodostaessa tulee siis tietää, että illatiivissa tai possessiivisuffiksillisessa muodossa ei kuulu käyttää astevaihtelua. Muita

huomionarvoisia muotoja otannassa tällä tasolla on myös muutamia, jotka ovat *Helsingissa*, *Helsingista*, *kaupungissä*, *rouan* ja *poikata*. Näissä kolmessa ensimmäisessä on valittu vokaalisoinnun vastainen muoto, muodossa *rouan* on vokaalit väärässä järjestyksessä, ja muodossa *poikata* on väärä partitiivin päätte. Nämä kaikki on kuitenkin luokiteltu norminmukaisiksi muodoiksi, koska niissä on käytetty sijamuodon kannalta astevaihtelua oikein.

Ensimmäisellä keskitason taitotasolla eli tasolla B1 jakauma on taas hieman laajempi kuin aiemmalla taitotasolla. Tällä tasolla on aiempien taitotasojen lailla eniten sanoja partitiivissa (40 %). Melkein yhtä paljon on sekä genetiivi- (17 %) että inessiivisijaisia (18 %) sanoja. Tämän jälkeen eniten on sanoja illatiivissa (8 %), elatiivissa (6 %), adessiivissa (6 %) ja allatiivissa (2 %). 1 % sanoista on sekä ablatiivissa että genetiivi + possessiivisuffiksi - ja allatiivi + possessiivisuffiksi -muodoissa. Ensiesiintymisensä tekevät siis allatiivi ja allatiivin possessiivisuffiksillinen muoto. Translatiiveja ei kuitenkaan ole tällä tasolla ollenkaan. Norminmukaisuusprosentti on tällä tasolla siis 94 %. Muutoin kaikissa sijamuodoissa ja suffiksillisissa muodoissa kaikki sanat ovat oikein, mutta partitiiveista on oikein 97,5 %, illatiiveista 87,5 %, elatiiveista 83 % ja genetiiveistä 82 %. Huomionarvoista on siis se, että ensiesiintymisensä tehneet muodot ovat molemmat oikein, vaikkakin niitä on yhteensä vain kolme. Adessiivi- ja inessiivimuotoiset sanat ovat kaikki ensimmäistä kertaa oikein tällä tasolla. Tällä tasolla on myös ensimmäistä kertaa mukana virheellinen partitiivimuoto, koska aiemmilla tasoilla partitiivi on ollut 100 % oikein.

Virheellisiä muotoja on tällä tasolla vain kuusi. Ne ovat genetiivissä *sikan*, *Kaurismäkin* ja *poikan*, partitiivissa *sängyä*, elatiivissa *jälkestä* ja illatiivissa *Helsingiin*. Elatiivisijainen norminvastainen muoto ja kaikki genetiivisijaiset norminvastaiset muodot ovat sellaisia, että niissä ei ole käytetty astevaihtelua, vaikka olisi pitänyt. Sekä partitiivi- että illatiivimuotoisessa norminvastaisessa muodossa on käytetty astevaihtelua, vaikka näin ei kuuluisi tehdä. Muita huomionarvoisia muotoja on tällä tasolla kaksi, jotka ovat molemmat partitiivissa, ja ne ovat *sänkyätä* ja *roukaa*. Ensimmäisessä on käytetty väärää partitiivin päätettä ja jälkimmäisessä on vokaalit väärässä järjestyksessä. Kummatkin ovat kuitenkin astevaihtelusääntöjen valossa oikein. Tällä tasolla sekä prosentuaalisesti että määrällisesti vaikein sijamuoto on genetiivi ja tyypillisin virhe on, että astevaihtelua ei käytetä, vaikka sitä pitäisi käyttää.

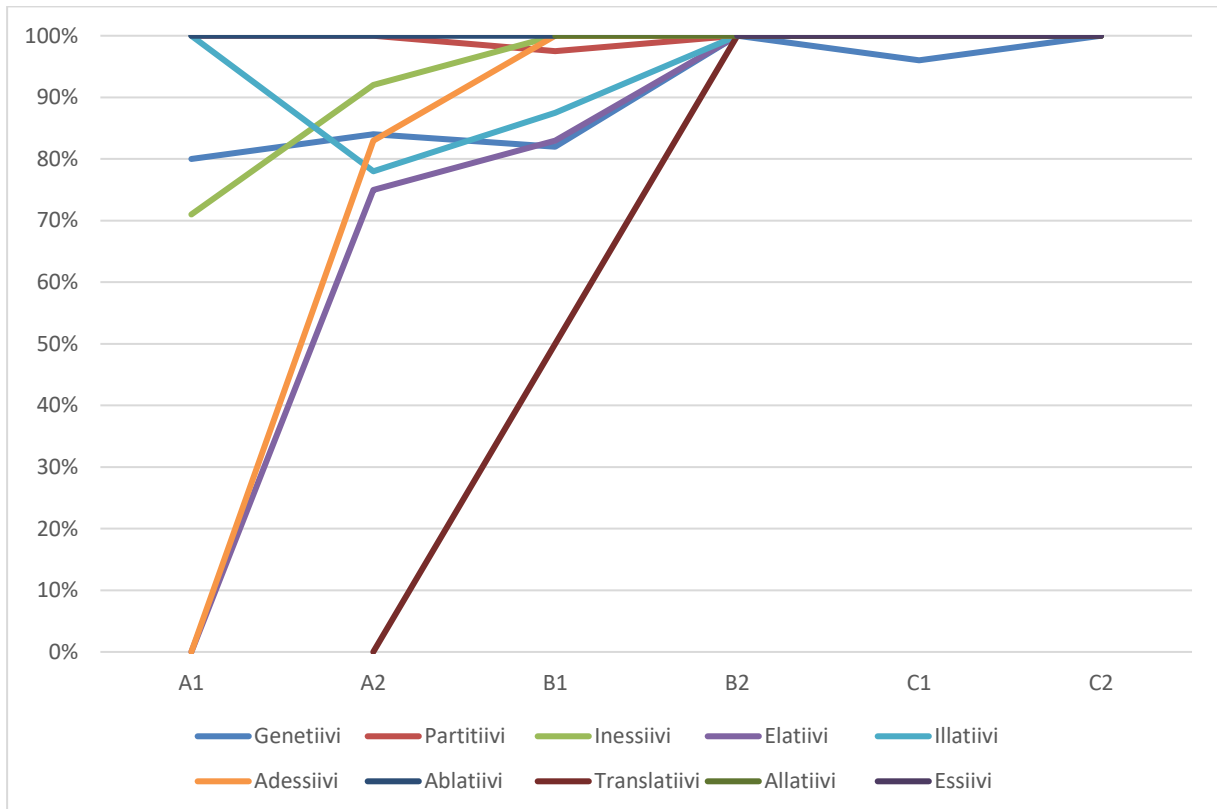
Taitotasolla B2 jakauma jakaantuu aika tasaisesti neljän sijamuodon kesken, jonka lisäksi muutamia sanoja on myös muissa muodoissa. Jakauma on seuraavanlainen: genetiivi 30 %, partitiivi 23 %, inessiivi 19 %, illatiivi 17 %, elatiivi 4 %, adessiivi 2 %, essiivi 2 %, partitiivi + possessiivisuffiksi -muoto 2 % ja translatiivi 1 %. Tällä tasolla esiintyy ensimmäisen kerran essiivi- ja partitiivi + possessiivisuffiksi -muoto. Tällä tasolla ei kuitenkaan ole yhtäkään ablatiivia, allatiivia tai genetiivi - ja allatiivi + possessiivisuffiksi -muotoa, vaikka näitä on esiintynyt jo aiemmilla tasoilla. Tällä tasolla kokonaisnorminmukaisuusprosentti on ensimmäistä kertaa 100, eli tällä tasolla ei ole yhtäkään norminvastaista muotoa. Yksi huomiota herättänyt muoto tällä tasolla kuitenkin on, joka on *Helsingissa*. Tässä on siis valittu vokaalisointua rikkova pääte, mutta astevaihtelua on käytetty oikein, joten tämä on luokiteltu norminmukaiseksi.

Jakauma on tasolla C1 taas vähän tasaisempi kuin aiemmin. 26 % sanoista on genetiivissä, 24 % partitiivissa, 22 % inessiivissä, 11 % sekä elatiivissa että illatiivissa, 2 % essiivissä, 1 % adessiivissa, allatiivissa sekä partiivi + possessiivisuffiksi ja adessiivi + possessiivisuffiksi -muodoissa. Ensimmäisen kerran esiintyy adessiivi + possessiivisuffiksi -muotoinen sana. Tällä tasolla ei kuitenkaan ole aiemmin esiintyneistä muodoista ablatiivia, translatiivia tai possessiivisuffiksillisia muotoja genetiivistä tai allatiivista. Norminmukaisuusprosentti on tällä tasolla 99, eli tällä tasolla on yksi norminvastainen muoto. Tämä on genetiivissä ja muoto on *Helsinki*. Tässä norminvastaisen muodon aiheuttaa se, että astevaihtelua ei ole käytetty.

Korkeimmalla tasolla eli tasolla C2 jakauma on taas hieman erilainen. Genetiivimuotoisia sanoja on 35 %, partitiivimuotoisia sanoja on 15 %, illatiivissa sanoja on 14 %, inessiivissä 11 %, elatiivissa ja allatiivissa sanoja on 7 %, partitiivi + possessiivisuffiksi -muotoja on 5 %, essiivissä sanoja on 3 % ja 1 % sanoista on translatiivissa ja elatiivi + possessiivisuffiksi -muodossa ja lisäksi myös partitiivi + *kaan*-partikkeli -muodossa. Tällä tasolla ei ole laisinkaan aiemmilla tasoilla esiintyneitä adessiivi - tai ablatiivimuotoisia sanoja, eikä possessiivisuffiksillisia muotoja, jotka olisivat genetiivissä, allatiivissa tai adessiivissa. Viimeisellä tasolla norminmukaisuusprosentti on taas 100, joten norminvastaisia muotoja ei ole laisinkaan. Yksi poikkeuksellinen muoto tasolla kuitenkin on, ja se on *iässä*, mutta kyseessä on mitä luultavimmin vain kirjoitus- tai lyöntivirhe.

Taitotasokohtainen erilaisten muotojen kokonaisjakauma on sellainen, että tasoilla A1 ja A2 on kahdeksan eri sijassa olevaa sanaa, B2 niitä on yhdeksän, B1 ja C1 näitä on kymmenen ja

C2-tasolla hajonta on suurin, koska eri sijamuotoja tai muita muotoja on 11. Eri muotojen jakauma siis nousee lineaarisesti tasolta toiselle muuten, mutta tasolla B1 on enemmän muotoja kuin tasolla B2.



Kuvio 3. Kvalitatiivisen vaihtelun sanojen sijamuotojen norminmukaisuuden kehitys taitotasoin. Sijamuotojen norminmukaisuusprosentit vaihtelevat hieman taitotasolta toiselle siirryttäessä. Kuvioista 3 voi nähdä, miten eri sijamuodoissa olevien sanojen norminmukaisuus muuttuu eri taitotasolla. Kuvioista on jätetty selkeyden vuoksi possessiivisuffiksilliset ja liitepartikkelin sisältävät muodot. Esimerkiksi adessiivi ja elatiivi nousevat heti ensiesiintymisensä jälkeen 0 %:sta nopeasti 80 %:n rajan tuntumaan. Inessiivi kehittyy melko lineaarisesti paremmaksi ensiesiintymisensä jälkeen. Illatiivin kehityskäyrä on mielenkiintoinen, koska se laskee 100 %:sta melko radikaalisti alle 80 %:iin, josta se nousee tasaisesti ja on taas 100 % taitotasolla B2. Genetiivi pysyy kolmella ensimmäisellä tasolla 80 %:n tuntumassa, jonka jälkeen se nousee lähelle 100 %. Partitiivi on koko ajan lähellä 100 %, mutta tasolla B1 se tekee pienen notkahduksen. Ablatiivi näkyy kuviossa huonosti, mutta se on koko sen esiintymisen ajan 100 %:ssa. Translatiivista ei ole jokaisella taitotasolla, kuten ei kaikkia muitakaan sijamuotoja, mutta esiintymisensä jälkeen se nousee pian 0 %:sta 100 %:iin. Sekä allatiivi että essiivi ovat koko esiintymisensä ajan 100 %:ssa, jonka takia ne näkyvät kuviossa 3 huonosti.

Tarkkuusprosentit muotokohtaisesti kertovat niiden haastavuudesta, tosin varauksella, koska tähän vaikuttaa radikaalisti esiintymien määrä. 100 % oikein jokaisella taitotasolla osattiin muodostaa ablatiivi, allatiivi, essiivi sekä allatiivi -, partitiivi -, adessiivi - ja elatiivi + possessiivisuffiksi -muoto sekä liitepartikkelin sisältävä partitiivimuoto. 99 %:sti oikein osattiin muodostaa partitiivi, 96 % inessiivi, 95 % illatiivi, 93 % genetiivi, 90 % elatiivi, 80 % genetiivi + possessiivisuffiksi, 78 % adessiivi ja 50 % translatiivi. Näin ollen siis translatiivi on ollut oppijoille haastavin muoto muodostaa. 100 %:sti norminmukaiset muodot ovat kaikki sellaisia, että niitä esiintyy ensimmäisen kerran vasta myöhemmillä taitotasoilla, paitsi ablatiivia esiintyy jo tasolla A1. Näiden myöhäinen esiintyminen ja pieni lukumäärä voisi selittää niiden korkean norminmukaisuusprosentin, koska myöhemmillä tasoilla oppijat ovat olleet tekemisissä astevaihtelusäännön kanssa pidempään, jolloin säännön soveltaminen on jo tutumpaa. Partitiivimuotojen suuri määrä ja korkea norminmukaisuusprosentti taas kertoo luotettavammin siitä, että sitä osataan käyttää astevaihtelun näkökulmasta todella hyvin. Tähänkin tosin vaikuttaa luonnollisesti se, että partitiivissa ei ole astevaihtelua. Sellainen sijamuoto, joita otannassa on paljon, ja jossa on korkea norminmukaisuusprosentti, ja jossa on astevaihtelua, on inessiivi. Inessiiviäkin osataan käyttää siis todella hyvin kaikkien taitotasojen läpi. Jos hallinnan rajana pidetään tuota monesti mainittua 80 %:n rajaa, vain translatiivi ja adessiivi eivät kokonaisprosentiltaan ole hallinnassa. Taitotasokohtaisesti jotkin muodot kuitenkin myös ylittävät 80 %-rajan, ja nämä ovat tasolla A1 inessiivi, elatiivi ja adessiivi ja tasolla A2 elatiivi, illatiivi, translatiivi ja genetiivi + possessiivisuffiksi. Kokonaisuudessaan millään taitotasolla oppijoiden kokonaisnorminmukaisuusprosentti ei ole alle 80 %.

#### 4.1.4 Sanojen pituuden jakauma ja tarkkuus

Kuten olen aiemmin maininnut, tässä tutkielmassa sanojen pituutta tarkastellaan tavujen määrän kautta. Tavut ovat siinä mielessä luonnollinen tarkastelukohde, että astevaihtelun pääsääntö nojaa juuri tavuihin ja niiden rakenteeseen. Kuten taulukossa 3 on nähtävissä, suurin osa otannan astevaihtelusanoista ovat kaksitavuisia (49 %), sen jälkeen vähän vähemmän on kolmetavuisia (38 %), toiseksi vähiten neljätavuisia (14 %) ja kaikkein vähiten viisitavuisia (0,2 %). Tavumäärän mukaiset tarkkuusprosentit ovat sellaisia, että viisitavuisista sanoista oikein on 100 %, kaksitavuisista 98 %, neljätavuisista 93 % ja kolmetavuisista sanoista oikein on 92 %. Näin ollen kolmetavuiset sanat ovat oppijoille haastavimpia.

Taulukko 3. Kvalitatiivisen vaihtelun sanojen pituuden jakauma ja tarkkuus.

	A1		A2		B1		B2		C1		C2		Yhteensä pituuksia	
	n=40		n=100		n=100		n=100		n=100		n=100			
	mää- rä	NM	määrä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	määrä	NM
<b>2 tavua</b>	50 % (20)	<b>95</b> % (19)	42 % (42)	<b>95</b> % (40)	50 % (50)	<b>96</b> % (48)	53 % (53)	<b>100</b> % (53)	45 % (45)	<b>100</b> % (45)	53 % (53)	<b>100</b> % (53)	49 % (263)	<b>98 %</b> (258)
<b>3 tavua</b>	35 % (14)	<b>79</b> % (11)	37 % (37)	<b>73</b> % (27)	33 % (33)	<b>91</b> % (30)	39 % (39)	<b>100</b> % (39)	43 % (43)	<b>98</b> % (42)	37 % (37)	<b>100</b> % (37)	38 % (203)	<b>92 %</b> (186)
<b>4 tavua</b>	15 % (6)	<b>50</b> % (3)	21 % (21)	<b>95</b> % (20)	17 % (17)	<b>94</b> % (16)	7 % (7)	<b>100</b> % (7)	12 % (12)	<b>100</b> % (12)	10 % (10)	<b>100</b> % (10)	14 % (73)	<b>93 %</b> (68)
<b>5 tavua</b>	-	-	-	-	-	-	1 % (1)	<b>100</b> % (1)	-	-	-	-	0,2 % (1)	<b>100</b> % (1)
<b>Yh- teen- sä</b>		<b>82,5</b> % (33)		<b>87</b> % (87)		<b>94</b> % (94)		<b>100</b> % (100)		<b>99</b> % (99)		<b>100</b> % (100)		

Taitotasolla A1 50 % sanoista on kaksitavuisia, 35 % kolmetavuisia ja loput 15 % ovat neljätavuisia. Vähiten norminmukaisia muotoja on tällä taitotasolla neljätavuisissa sanoissa (50 %), toiseksi vähiten kolmetavuisissa (79 %) ja eniten kaksitavuisissa (95 %) sanoissa. Näin ollen ainakin tällä taitotasolla voi todeta, että pidemmän sanan taivuttaminen on oppijoille haastavampaa. Tosin määrällisesti sekä kolme- että neljätavuisissa sanoissa on molemmassa kolme norminvastaista muotoa. Kaksitavuisissa sanoissa virheellinen muoto on *sänkyn*, kolmetavuisissa sanoissa muodot ovat *sänkyllä*, *sänkyllä* ja *sänkyssä* ja neljätavuisissa sanoissa *aurinkosta*, *kaupunissa* ja *kauppunkilla*. Kaikissa muissa virheellisissä muodoissa päätteän lisääminen lisää sanan yhden tavun, paitsi kaksitavuiseseen muotoon *sänky*, jossa tavumäärä pysyy samana. Kaikissa muissa kuin kaksitavuisessa sanassa päätte on muodossa CCV, ja kaksitavuisessa sanassa päätte on yksi konsonantti. Oppijat eivät tällä tasolla ylittä 80 % -norminmukaisuusrajaa kaksi- ja kolmitavuisissa sanoissa.

A2-tasolla kaksitavuisia sanoja on 42 %, kolmetavuisia sanoja on 37 % ja neljätavuisia 21 %. Kaksi- ja kolmetavuisia sanoja on siis prosentuaalisesti melkein yhtä paljon. Edellisestä taitotasosta poiketen norminmukaisuus ei kuitenkaan laske pidemmissä sanoissa. Tällä tasolla kaksitavuisista sanoista 95 % on norminmukaisia, kolmetavuisista 73 % ja neljätavuisista taas 95 %. Kaksitavuisista sanoista norminvastaisia muotoja ovat *sikan* ja *sänkyn*, kolmetavuisista *ruokaksi*, *ruokaksi*, *alunsa*, *Turkussa*, *kaupungiin*, *kaupungiin*, *lenikin*, *helsinkin*, *henkestä* sekä *sänkyllä* ja neljätavuihin norminvastainen muoto on *aurinkossa*. Tavumäärät pysyvät

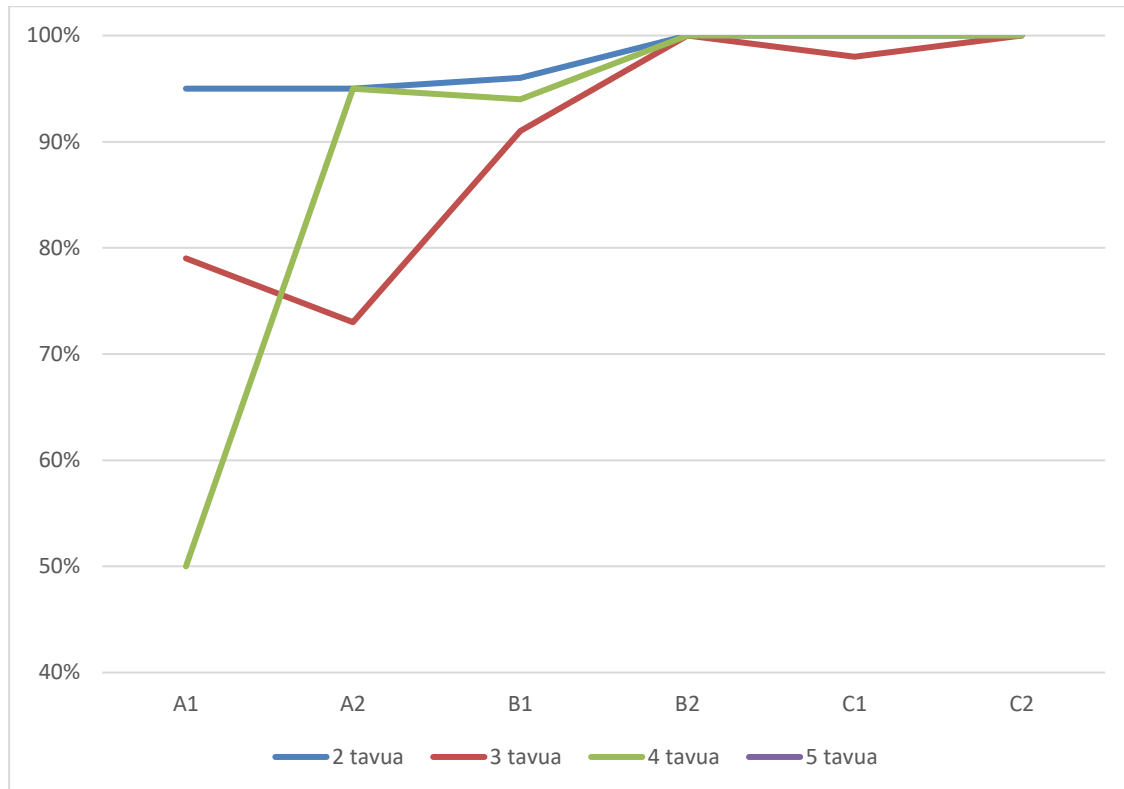
pääteestä huolimatta samana muodoissa *sikan, sänkyn, kaupungiin, lenikin ja helsinkin*. Lopuissa norminvastaisissa muodoissa tavumäärä lisääntyy yhdellä. Niihin sanoihin, joissa tavumäärä lisääntyy, lisätään CCV-muotoinen tavu. Sanoihin, joissa tavumäärä pysyy samana, lisätään vain yksi konsonantti, paitsi muotoon *kaupungiin*, jossa päätteenä on CV. Tällä tasolla oppijoille haastavimpia ovat siis olleet kolmetavuiset sanat, joihin lisätään CCV-muotoinen pääte. A2-tasolla hallinnassa eivät vielä 80 %:n rajaan suhteutettuna ole sanat, joissa on kolme tavua.

Seuraavalla taitotasolla B1 sanoista 50 % on kaksitavuisia, 33 % kolmetavuisia ja 17 % neljätavuisia. Tällä tasolla on siis eniten lyhyitä sanoja ja määrä vähenee tavumäärän kasvaessa. Tälläkään tasolla norminmukaisuus ei kuitenkaan kulje linjassa jakauman kanssa, vaan sen sijaan 96 % kaksitavuisista sanoista on norminmukaisia, kolmetavuisista sanoista 91 % on norminmukaisia ja neljätavuisista sanoista 94 % on norminmukaisia. Kaksitavuisten sanojen norminvastaiset muodot ovat *sikan* sekä *poikan*, kolmetavuisten *jälkestä*, *Helsinkiin* ja *sängyä*, ja neljätavuisten sanojen norminvastainen muoto on *Kaurismäkin*. Kaikissa muissa muodoissa tavumäärä pysyy samana, mutta muodoissa *jälkestä* ja *sängyä* tavumäärä lisääntyy yhdellä. Tavumäärältään samana pysyviin sanoihin lisätään joko yksi konsonantti tai CV. Muotoon *jälkestä* on lisätty tavu, joka on muotoa CCV. *sängyä*-muotoon on lisätty yksi vokaali, jolloin tässä sanassa ei kuuluisi olla astevaihtelua, mutta sitä on kuitenkin käytetty. Tällä tasolla oppijoille haastavimpia ovat kolmetavuiset sanat, mutta lisättävän päätteen kannalta vaikein lisäys on yksi konsonantti, koska tällaisissa muodoissa virheitä on kolmessa eri sanassa.

Tasolla B2 jakauma on laajin, koska ainoastaan tällä tasolla on mukana yksi viisitavuinen sana. Suurin osa sanoista on kuitenkin kaksi- ja kolmetavuisia, joita on 53 % ja 39 %. Neljätavuisia sanoja on vain 7 % ja viisitavuisia on 1 %. Tällä tasolla ei ole yhtäkään norminvastaista muotoa, joten sanan pituus ei ainakaan tässä vaiheessa vaikuta astevaihtelun norminmukaisuuteen. Tasolla C1 jakauma on hyvin samanlainen kuin aiemmallakin tasolla. Kaksitavuisia sanoja on 45 %, kolmetavuisia on 43 % ja neljätavuisia on 12 %.

Norminmukaisuusprosentti on muutoin 100, mutta kolmetavuisissa sanoissa on yksi virheellinen muoto, joka laskee norminmukaisuusprosentin 98:aan. Tämä kyseinen muoto on *Helsinkiin*, jolloin lisättävä pääte ei nosta tavumäärää, ja pääte on vain yksi konsonantti. Näin ollen voidaan sanoa, että tällä tasolla haastavinta on ollut taivuttaa kolmetavuista sanaa niin, että päätteenä on vain yksi konsonantti. Tällöin päätteen lisäys muuttaa viimeisen tavun CV-muodosta muotoon CVC. Myös tasolla C2 jakauma on hyvin samanlainen kuin aiemmin.

Tällä tasolla kaksitavuisia sanoja on 53 %, kolmetavuisia sanoja on 37 % ja neljätavuisia 10 %. Tälläkään tasolla ei ole yhtäkään astevaihtelun kannalta virheellistä muotoa.



Kuvio 4. Kvalitatiivisen vaihtelun sanojen norminmukaisuuden kehitys tavujen määrän mukaan taitotasoin.

Tietyn tavumäärän sisältävien sanojen norminmukaisuusprosentti vaihtelee hyvin eri tavoin. Kuvio 4 voi nähdä, miten kaksitavuiset sanat osataan suhteellisen hyvin jo heti alussa, jonka jälkeen norminmukaisuusprosentti kasvaa taitotasolta toiselle edetessä. Kolmetavuisien sanojen norminmukaisuus ensin laskee merkittävästi, jonka jälkeen se nousee jyrkästi ylöspäin ja pysyy 100 %:n tuntumassa. Neljätavuiset sanat osataan selvästi huonoiten ensimmäisellä taitotasolla, jonka jälkeen norminmukaisuus nousee hyvin jyrkästi jo seuraavalla taitotasolla, josta se nousee pienen notkahduksen kautta 100 %:iin. Viisitavuisia sanoja on vain yhdellä tasolla, jonka takia tätä ei näy kuviossa 4, mutta norminmukaisuus on 100 %, joten kuviossa 4 sen pitäisi näkyä yhtenä pisteenä taitotason B2 kohdalla.

#### 4.1.5 Sanatyyppien jakauma ja tarkkuus

Sanatyyppit ovat astevaihtelun kannalta mielenkiintoisia, koska kaikissa sanatyypeissä ei esiinny astevaihtelua laisinkaan. Tutkielman rajauksen puitteissa astevaihtelua esiintyy kolmessa eri sanatyypissä. Mielenkiintoista on, että koko otannassa ei ollut yhtäkään sellaista

sanaa, jotka eivät kuuluisi näihin sanatyyppeihin, joten oppijat eivät ole soveltaneet ainakaan *k:n* astevaihtelua sellaisiin sanoihin, jossa sitä ei sanatyypin puitteissa voisi olla. Tosin sanatyyppi määräytyy sanan rakenteen mukaan, jolloin *k:n* astevaihtelun käyttäminen joissain toisissa kuin näissä sanatyypeissä vaatisi aika radikaaleja muutoksia sanan rakenteeseen eikä sana olisi välttämättä edes tunnistettavissa. Kuten taulukosta 4 näkyy, koko otannassa selkeästi eniten kvalitatiivisen vaihtelun sanoista ovat sanatyypin KATTO sanoja (72 %). Toiseksi eniten on PANKKI-tyyppin sanoja (22 %) ja vähiten LEHTI-tyyppin sanoja (6 %). Sanatyypikohtaista tarkkuutta tarkasteltaessa huomataan, että KATTO-tyyppin sanoista oikein on 96 %, PANKKI-tyyppin sanoista 93 % ja LEHTI-tyyppin sanoista oikein on 91 %. Näin ollen voi sanoa, että prosentuaalisesti oppijoille haastavin sanatyypin astevaihtelun kannalta on LEHTI-sanatyyppi.

Taulukko 4. Kvalitatiivisen vaihtelun sanojen sanatyypin jakauma ja tarkkuus.

	A1		A2		B1		B2		C1		C2		Yhteensä sanatyyppejä	
	n=40		n=100		n=100		n=100		n=100		n=100			
	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM
<b>KAT- TO</b>	72,5 % (29)	<b>83</b> % (24)	63 % (63)	<b>87</b> % (55)	72 % (72)	<b>96</b> % (69)	72 % (72)	<b>100</b> % (72)	79 % (79)	<b>100</b> % (79)	72 % (72)	<b>100</b> % (72)	72 % (38 7)	<b>96 %</b> (371)
<b>PANK- KI</b>	25 % (10)	<b>80</b> % (8)	35 % (35)	<b>89</b> % (31)	22 % (22)	<b>95</b> % (21)	20 % (20)	<b>100</b> % (20)	16 % (16)	<b>94</b> % (15)	16 % (16)	<b>100</b> % (16)	22 % (11 9)	<b>93 %</b> (111)
<b>LEH- TI</b>	2,5 % (1)	<b>100</b> % (1)	2 % (2)	<b>50</b> % (1)	6 % (6)	<b>66</b> % (4)	8 % (8)	<b>100</b> % (8)	5 % (5)	<b>100</b> % (5)	12 % (12)	<b>100</b> % (12)	6 % (34)	<b>91 %</b> (31)
<b>Yh- teen- sä</b>		<b>82,5</b> % (33)		<b>87</b> % (87)		<b>94</b> % (94)		<b>100</b> % (10 0)		<b>99</b> % (99)		<b>100</b> % (100)		

Jakaumat ovat jokaisella taitotasolla hyvin samanlaiset. Taitotasolla A1 KATTO-sanatyypin sanoja on 72,5 % sanoista, PANKKI-tyyppin sanoja on 25 % ja LEHTI-tyyppin sanoja on 2,5 %. KATTO-sanojen normimukaisuusprosentti on 83 %, PANKKI-sanojen 80 % ja LEHTI-sanojen 100 %. Tosin viimeiseksi mainitun tyyppin sanoja on vain yksi. KATTO-sanatyypin virheelliset muodot ovat *sänkyllä, sänkyllä, aurinkosta, sänkyssä ja sänkyyn*. Tämän tyyppin sanat ovat tavallisimpia suomen kielen sanastossa. Tämän tyyppin taivutuksessa ei tarvitse kiinnittää huomiota mihinkään muuhun, kuin astevaihteluun päätettä lisätessä. PANKKI-tyyppin sanoista virheelliset muodot ovat *kaupunissa* ja *kauppunkilla*. Myöskään tässä tyyppissä

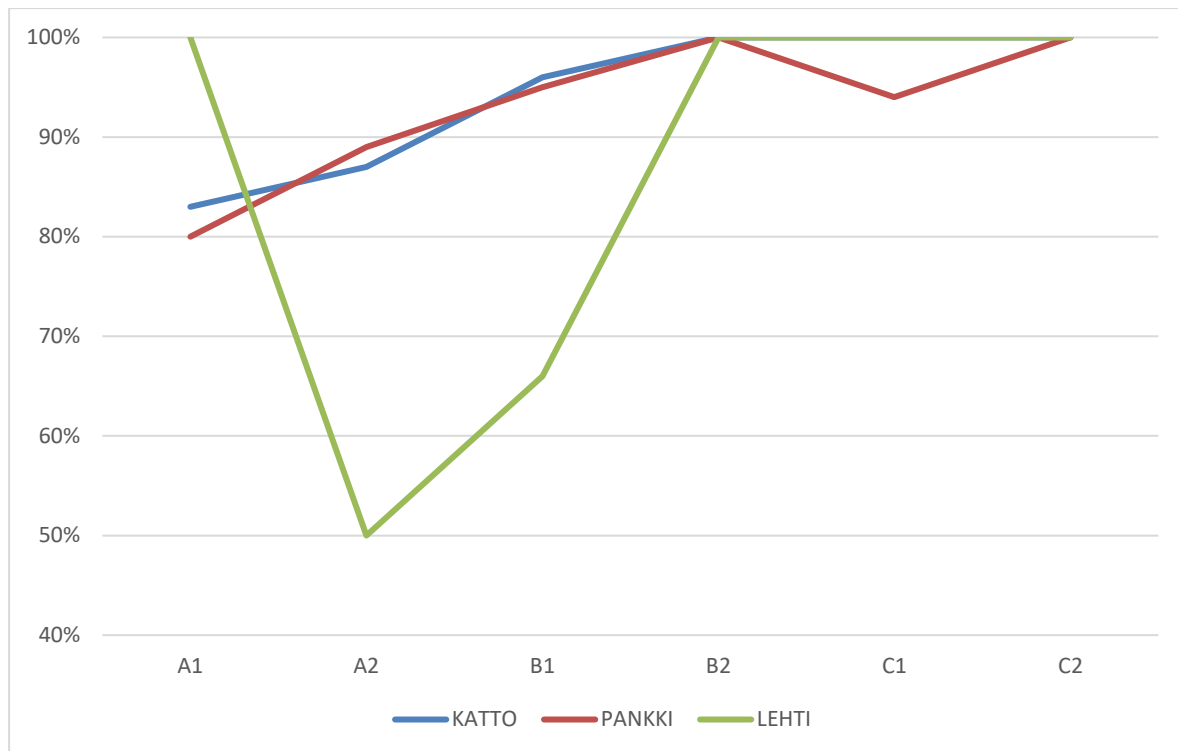
ei taivuttaessa tarvitse kiinnittää huomiota muuhun kuin astevaihteluun, koska tässä sanatyypissä ei ole vokaalivaihteluita. Siinä mielessä tosin tämän sanatyypin sanat voivat olla vaikeampia muodostaa kuin esimerkiksi KATTO-sanat, koska *i*-loppuisista sanoista ei automaattisesti voi tietää mihin *i*-loppuiseen sanatyyppiin sana kuuluu. Nämä sanat on kuitenkin osattu taivuttaa oikean sanatyypin mukaisesti, koska näissä muodoissa ei ole tarpeettomasti tehty vokaalivaihteluita. Kumpikin virheellinen muoto on muiltakin osin norminvastainen, koska muodossa *kaupunissa* ei ole käytetty *nk : ng* -vaihtelua, vaan *k* on jätetty pois, kuten jo vaihtelupariluvussa 5.1.2 totesin. Muodossa *kauppunkilla* astevaihtelua ei ole käytetty, vaan sanassa on edelleen vahva vaihtelupari. Sanassa on myös kaksi *p*-kirjainta, vaikka pitäisi olla vain yksi. Tällä tasolla on myös yksi muuten huomiota herättävä sanamuoto, joka on *väkensä*. Tässä astevaihtelua on käytetty ihan oikein, koska sanassa on possessiivisuffiksi, mutta sanaa ei ole taivutettu sen sanatyypin mukaisesti, jota se edustaa. Oikea taivutettu muoto olisi *väkensä*, jolloin voidaan ajatella, että oppija on joko unohtanut tehdä *i-e*-vaihdoksen tai sitten oppija on tulkinnut sanan olevan PANKKI-sanatyyppiä, vaikka se kuuluu LEHTI-sanatyyppiin, jossa tätä kyseistä vaihtelua on. Prosentuaalisesti tällä tasolla haastavin sanatyyppi astevaihtelun kannalta on PANKKI.

Tasolla A2 jakauma on seuraavanlainen: KATTO-sanatyypin sanoja on 63 %, PANKKI-tyypin sanoja on 35 % ja LEHTI-sanatyypin sanoja on 2 %. Norminmukaisuus on tällä tasolla KATTO-sanoissa 87 %, PANKKI-sanoissa 89 % ja LEHTI-sanoissa se on 50 %. Tosin LEHTI-sanoja on tällä tasolla vain kaksi, joista toinen on norminvastainen. KATTO-tyypin norminvastaiset muodot ovat *ruokaksi*, *ruokaksi*, *alunsa*, *Turkussa*, *sikan*, *aurinkossa*, *sänkyn* ja *sänkyllä*. Tässä tyypissä ei siis tarvitse kiinnittää taivuttaessa muuhun kuin astevaihteluun huomiota. PANKKI-tyypin sanoista norminvastaisia muotoja ovat *kaupungiin*, *kaupungiin*, *lenikin* ja *helsinkin* ja LEHTI-tyypin norminvastainen muoto on *henkestä*. Näitä molempia sanatyyppejä yhdistää se, että sanan sanatyyppi täytyy tietää, jotta osaa muodostaa oikein vartalon vokaalivaihteluineen. Näissä muodoissa ei sen suhteen ole ongelmia, ja LEHTI-tyypin sanassakin on osattu tehdä oikea vokaalivaihtelu. Norminvastaisuus syntyy siitä, että kaikissa muodoissa on sijamuotoon suhteutettuna vääränvahvuinen vaihtelupari. Sanasta *lenikin* tosin puuttuu myös yksi kirjain, jolloin luonnollisesti sitä ei ole osattu taivuttaa *nk : ng* -vaihteluparin mukaisesti. Tällä tasolla oppijoille on prosentuaalisesti ollut vaikeinta taivuttaa LEHTI-sanatyypin sanoja.

Taitotasolla B1 jakauma on sellainen, että KATTO-tyypin sanoja on 72 %, PANKKI-sanoja on 22 % ja LEHTI-sanoja on 6 %. Jos tätä jakaumaa vertaa edellisen taitotason jakaumaan,

tällä tasolla on merkittävästi enemmän KATTO-sanoja kuin PANKKI-sanoja. Myös LEHTI-sanojen määrä on kasvanut. Norminmukaisuus on tällä tasolla merkittävästi korkeampi sanatyypeittäin kuin aiemmalla tasolla. KATTO-sanoista norminmukaisia muotoja on 96 %, PANKKI-sanoista 95 % ja LEHTI-sanoista 66 %. KATTO-sanoista norminvastaisia muotoja ovat *sikan*, *poikan* ja *sängyä*. PANKKI-sanoista virheellinen muoto on *Helsingiin* ja LEHTI-sanojen virheelliset muodot ovat *jälkestä* ja *Kaurismäkin*. Melkein kaikkia näitä muotoja on taivutettu sanatyypin valossa oikein, mutta kaikissa näissä on väärä vaihtelupari suhteessa sijamuotoon. *Kaurismäkin*-muoto on virheellinen myös sanatyypin näkökulmasta, koska sana *mäki* kuuluu LEHTI-sanatyyppiin, jolloin tässä sanassa kuuluisi olla *i-e*-vokaalivaihtelua. Tässä tapauksessa kyse on kuitenkin sukunimestä, jolloin oppija on voinut ajatella, että se ei taivu kyseisen sanatyypin mukaisesti. Sukunimet kuitenkin taipuvat samoin kuin samanmuotoinen yleissanakin (Kielitoimiston ohjepankki, Nimien taivutus: sukunimet, 2024, sähköinen lähde). Myös tällä tasolla oppijoilla on prosentuaalisesti ollut eniten haasteita LEHTI-sanatyypin taivutuksessa.

B2-tasolla jakauma on melkein täysin samanlainen kuin edellisellä tasolla. KATTO-sanatyypin sanoja on 72 %, PANKKI-tyypin sanoja on 20 % ja LEHTI-sanatyypin sanoja on 8 %. Tällä tasolla on siis taas hieman enemmän LEHTI-tyypin sanoja. Kuten aiemmin todettua, tällä tasolla kaikki sanat ovat norminmukaisia, joten tällä tasolla sanan sanatyypillä ei ole vaikutusta sen taivutuksen norminmukaisuuteen. C1-tasolla jakauma muuttuu taas hieman, ja tällä tasolla KATTO-sanoja on 79 %, PANKKI-sanoja 16 % ja LEHTI-sanoja 5 %. Tällä tasolla on siis enemmän KATTO-sanoja ja vähemmän LEHTI-sanoja. Norminmukaisuus on tällä tasolla KATTO- ja LEHTI-sanoissa 100 % ja PANKKI-sanoissa 94 %. Norminvastainen muoto on *Helsinkin*, ja kyseessä on sellainen sanatyyppi, jossa ei tapahdu vokaalivaihteluita. Näin ollen norminvastaisuuden aiheuttaa ainoastaan vahva-asteinen vaihtelupari. Tällä tasolla vaikein sanatyyppi on silloin PANKKI, koska sen sanatyypin sanat sisältävät yhden virheellisen muodon. Viimeisellä tasolla C2 jakauma on sellainen, että KATTO-sanatyypin sanoja on 72 %, PANKKI-tyypin sanoja on 16 % ja LEHTI-tyypin sanoja on 12 %. LEHTI-tyypin sanoja on siis korkeimmalla tasolla kaikkein eniten verrattuna muiden taitotasojen määriin. Tälläkin tasolla kaikki muodot ovat norminmukaisia, jolloin sanatyyppi ei vaikuta sanan taivutuksen vaikeuteen.



Kuvio 5. Kvalitatiivisen vaihtelun sanojen norminmukaisuuden kehitys sanatyypeittäin taitotasoittain. KATTO- ja PANKKI-sanatyypin norminmukaisuus kehittyy melko samalla tavalla suhteessa toisiinsa. LEHTI-sanatyypin norminmukaisuuden kehitys on hyvin erilainen, kuten kuviosta 5 näkee. Sekä KATTO- että PANKKI-tyypin sanat kehittyvät tasaisesti paremmiksi, kunnes ne tasolla B2 saavuttavat 100 %:n, tosin PANKKI-sanat tekevät vielä pienen notkahduksen tasolla C1. LEHTI-sanat sen sijaan lähtevät 100 %:sta, jonka jälkeen norminmukaisuus tippuu hyvin alas, jonka jälkeen se nousee jyrkästi paremmaksi ja saavuttaa 100 %:n taitotasolla B2.

#### 4.1.6 Sanojen syntaktisen funktion jakauma ja tarkkuus

Aiemmin tarkastelemani muuttujat ovat keskittyneet pitkälti sanojen rakenteeseen ja rakenteiden ominaisuuksiin. Tässä luvussa tarkastelen syntaktisen funktion vaikutusta käytettyihin astevaihtelumuotoihin, joka poikkeaa perustavanlaatuisesti aiemmista analyysiluvuista. Kuten olen aiemmin maininnut, sanat on tulkittu tietyn lauseenjäsenen edustajaksi riippumatta siitä, ovatko ne lauseenjäsenen edellyttämässä sijamuodossa. Sanoja on tulkittu myös tietyksi lauseenjäseneksi sen mukaan, minkälaisen lausekkeen osana ne ovat. Kuten taulukosta 5 näkyy, kaikilla taitotasoilla ei ole tarkasteltavia astevaihtelusanoja kaikissa syntaktisissa funktioissa.

Taulukko 5. Kvalitatiivisen vaihtelun sanojen syntaktisen funktion jakauma ja tarkkuus.

	A1		A2		B1		B2		C1		C2		Yhteensä syntaktista funktiota	
	n=40		n=100		n=100		n=100		n=100		n=100			
	Mää- rä	NM	määrä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	N M	Mää- rä	NM
<b>Sub- jekti</b>	2,5 % (1)	<b>100</b> % (1)	2 % (2)	<b>100</b> % (2)	3 % (3)	<b>100</b> % (3)	-	-	3 % (3)	<b>100</b> % (3)	6 % (6)	<b>100</b> % (6)	3 % (15)	<b>100</b> % (15)
<b>Ob- jekti</b>	32,5 % (13)	<b>100</b> % (13)	23 % (23)	<b>91</b> % (21)	26 % (26)	<b>92</b> % (24)	22 % (22)	<b>100</b> % (22)	19 % (19)	<b>100</b> % (19)	22 % (22)	<b>100</b> % (22)	23 % (125)	<b>97</b> % (121)
<b>Ad- ver- biaali</b>	52,5 % (21)	<b>67</b> % (14)	69 % (69)	<b>84</b> % (58)	52 % (52)	<b>96</b> % (50)	69 % (69)	<b>100</b> % (69)	74 % (74)	<b>99</b> % (73)	69 % (69)	<b>100</b> % (69)	65,5 % (354)	<b>62</b> % (333)
<b>Pred- ika- tiiv</b>	2,5 % (1)	<b>100</b> % (1)	-	-	2 % (2)	<b>50</b> % (1)	4 % (4)	<b>100</b> % (4)	-	-	2 % (2)	<b>100</b> % (2)	2 % (9)	<b>89</b> % (8)
<b>e- sub- jekti</b>	10 % (4)	<b>100</b> % (4)	6 % (6)	<b>100</b> % (6)	17 % (17)	<b>94</b> % (16)	5 % (5)	<b>100</b> % (5)	4 % (4)	<b>100</b> % (4)	1 % (1)	<b>100</b> % (1)	7 % (37)	<b>97</b> % (36)
<b>Yh- teen- sä</b>		<b>82,5</b> % (33)		<b>87</b> % (87)		<b>94</b> % (94)		<b>100</b> % (100)		<b>99</b> % (99)		<b>100</b> % (100)		

Taitotasolla A1 on astevaihtelusanoina kaikissa syntaktisissa funktioissa. Kaikkein eniten tällä tasolla on adverbiaaleja (52,5 %), sitten objekteja (32,5 %), sitten e-subjekteja (10 %) ja viimeiseksi subjekteja ja predikatiiveja (2,5 %). Muutoin kaikkien muiden syntaktisten funktioiden edustajien norminmukaisuus on 100 %, mutta adverbiaaleista oikein on vain 67 %. Näin ollen siis kaikki tämän tason virheelliset muodot ovat adverbiaaleja. Seuraavalla taitotasolla A2 on kaikkien muiden syntaktisten funktioiden edustajia paitsi predikatiiveja. Tälläkin tasolla on eniten adverbiaaleja (69 %), sitten objekteja (23 %), sitten e-subjekteja (6 %) ja viimeiseksi subjekteja (2 %). Kaikki subjektit ja e-subjektit ovat norminmukaisia, mutta objekteista oikeita muotoja on 91 % ja adverbiaaleista 84 %. Tälläkin tasolla haastavin lauseenjäsen on siis adverbiaali.

Taitotasolla B1 on taas kaikkien syntaktisten funktioiden edustajia. Näistä lähes puolet ovat adverbiaaleja (52 %), objekteja on 26 %, e-subjekteja on 17 %, subjekteja on 3 % ja predikatiiveja 2 %. Tällä tasolla norminmukaisuusprosentit vaihtelevat enemmän. Subjektit ovat kaikki norminmukaisia, adverbiaaleista norminmukaisia on 96 %, e-subjekteista 94 %,

objekteista 92 % ja predikatiiveista 50 %. Määrällisesti erot ovat kuitenkin hyvin pieniä, mutta prosentuaalisesti haastavin lauseenjäsien on kuitenkin predikatiivi. Määrällisesti sekä objekteja että adverbiaaleja on kuitenkin enemmän väärin, koska molemmissa on kaksi norminvastaista muotoa. Toisella perustason taitotasolla B2 on kaikkia muita lauseenjäsieniä paitsi subjekteja. Adverbiaaleja on taas eniten eli 69 %, objekteja on 22 %, 5 % on e-subjekteja, ja predikatiiveja on 4 %. Tällä tasolla kaikki oppijoiden käyttämät muodot ovat oikein astevaihtelun näkökulmasta.

Taitotasolla C1 on kaikkia muita lauseenjäsieniä paitsi predikatiiveja. Suurin osa sanoista on adverbiaaleja (74 %), objekteja on 19 %, e-subjekteja on 4 %, ja subjekteja on 3 %. Muutoin kaikki sanamuodot ovat oikein, mutta mukana on yksi virheellinen adverbiaalisana.

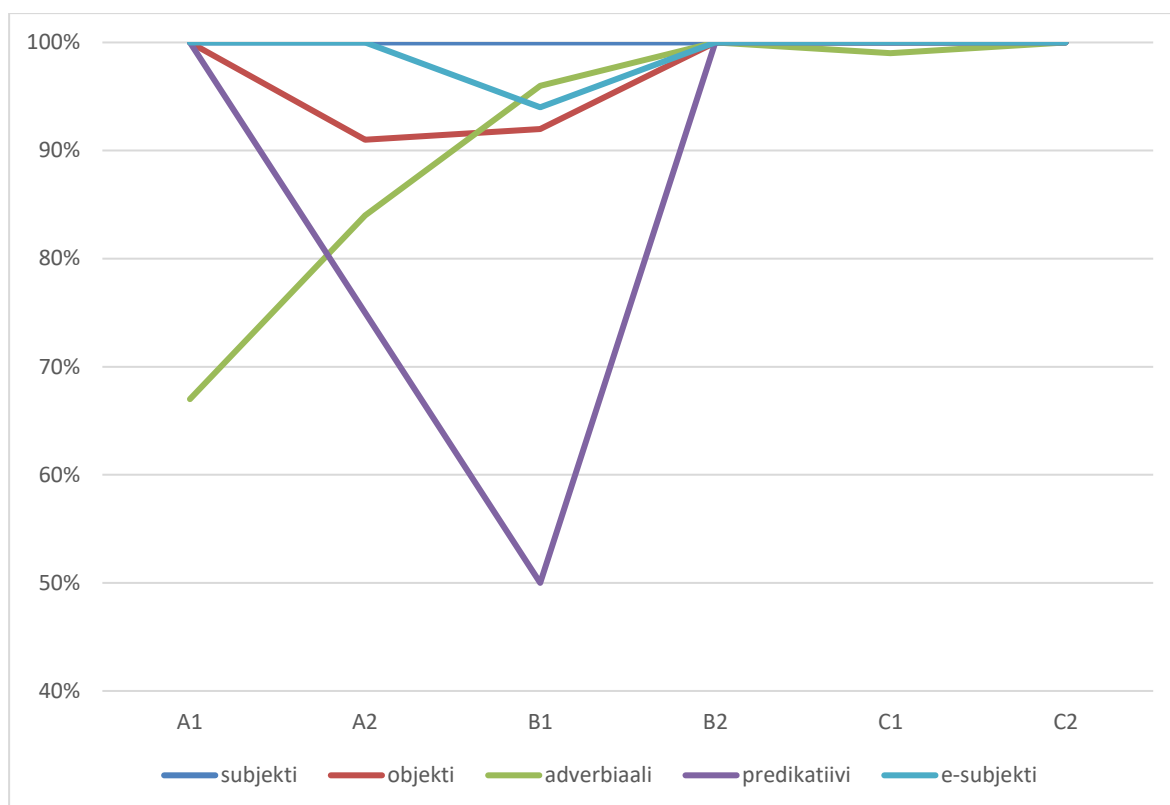
Korkeimmalla taitotasolla C2 on taas kaikkia lauseenjäsieniä. Jakauma on hyvin samanlainen kuin aiemminkin. Adverbiaaleja on 69 %, objekteja 22 %, subjekteja 6 %, predikatiiveja 2 % ja e-subjekteja 1 %. Tälläkin tasolla kaikki muodot ovat norminmukaisia.

Kaikista kvalitatiivisen vaihtelun otannan sanoista eniten on adverbiaaleja (65,5 %), sitten objekteja (23 %), sitten e-subjekteja (7 %), sitten subjekteja 3 %, ja predikatiiveja on 2 %. Kaikkien subjektien norminmukaisuusprosentti on 100, e-subjekteista ja objekteista norminmukaisia on 97 %, predikatiiveista 89 % ja adverbiaaleista 62 %. Kokonaisuudessaan siis adverbiaali on ollut oppijoille haastavin lauseenjäsien. Tämä selittyy ainakin osittain sillä, että subjektin sijamuodot ovat nominatiivi, genetiivi ja partitiivin, objektin sijat ovat nominatiivi, genetiivi, partitiivi ja akkusatiivi, predikatiivin sijat ovat nominatiivi, genetiivi ja partitiivi ja viimeisenä adverbiaalin sijat ovat inessiivi, elatiivi, illatiivi, adessiivi, ablatiivi, allatiivi, essiivi, translatiivi, komitatiivi, instruktiivi ja abessiivi, tosin myös genetiivi ja partitiivi voivat esiintyä adverbiaaleina, mutta rajallisesti (VISK § 1222 ja 1223).

Adverbiaalien mahdollisten sijamuotojen määrän takia näitä on otannassa eniten.

Adverbiaaleissa on myös eniten sellaisia sijamuotoja, joissa kuuluisi olla astevaihtelua.

Sijamuotokohtaisista norminmukaisuuksista olen kirjoittanut jo aiemmin luvussa 5.1.3.



Kuvio 6. Kvalitatiivisen vaihtelun sanojen norminmukaisuuden kehitys taitotasoinen tietyssä syntaktisessa funktiossa.

Kaikki subjektia edustavat otannan sanat ovat 100 %:sti oikein, joten ne näkyvät kuviossa 6 suorana viivana kuvion yläreunassa. Objektit ovat aluksi kokonaan oikein, jonka jälkeen ne ovat hetken lähellä 90 %, jonka jälkeen norminmukaisuus nousee taas 100 %:iin.

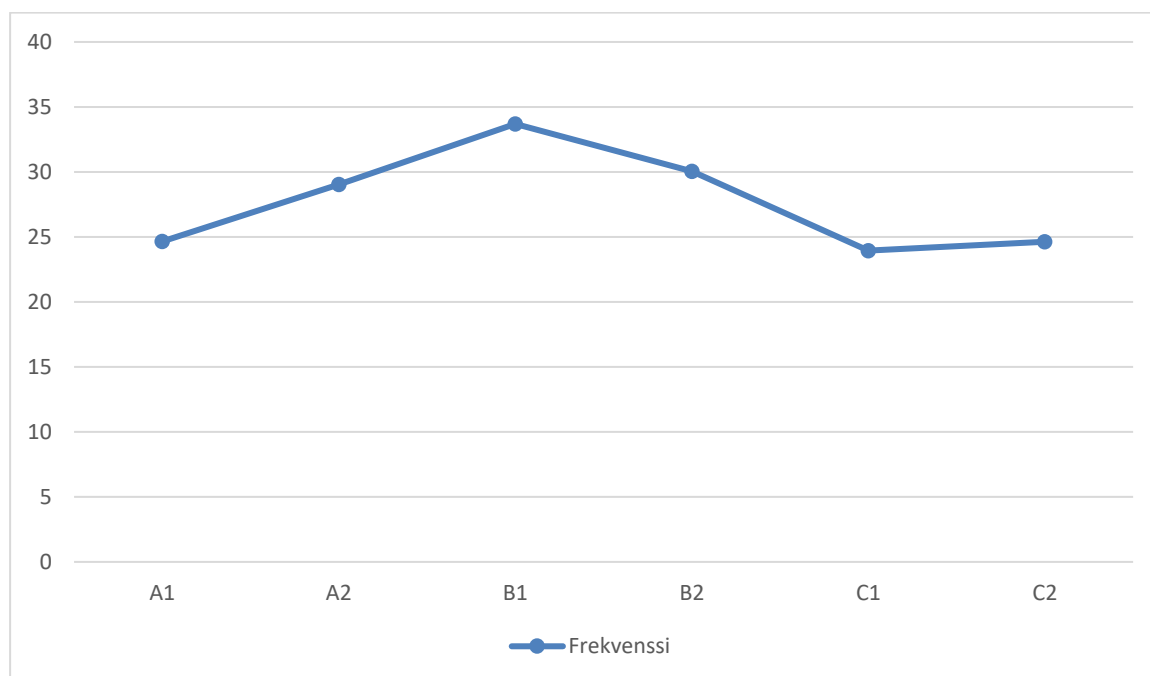
Adverbiaalien norminmukaisuus kehittyy suhteellisen tasaisesti taitotasojen A1–B1 välillä, jonka jälkeen norminmukaisuus on loppuilla taitotasolla 100 %. Predikatiiveja ei ole jokaisella taitotasolla, mutta niiden norminmukaisuus lähtee 100 %:sta, jonka jälkeen se tippuu radikaalisti 50 %:iin, jonka jälkeen se nousee taas radikaalisti 100 %:iin. E-subjektien norminmukaisuus on melkein jokaisella taitotasolla 100 %, mutta taitotasolla B1 se tekee pienen notkahduksen.

## 4.2 Kvantitatiivisen vaihtelun sanat

### 4.2.1 Frekvenssi

Luvussa 5.1.1 olen kirjoittanut frekvenssistä enemmän myös yleisemmällä tasolla, jonka takia en toista näitä samoja asioita tässä luvussa, koska ne koskevat osaltaan myös kvantitatiivisen vaihtelun alaisten sanojen frekvenssiä. Kuviosta 7 näkee, miten frekvenssi vaihtelee eri taitotasojen välillä. Tämänkin vaihtelun osalta frekvenssi on laskettu vertaamalla k:n

astevaihtelusanojen määrää kaikissa muissa sijamuodoissa kuin nominatiivissa kaikkien substantiivien määrään, joita kyseisellä taitotasolla on. Nämäkin frekvenssit on laskettu tuhatta sanaa kohden, jolloin taitotasojen sanojen määrät eivät vaikuta frekvenssiin niin paljoa.



Kuvio 7. Kvantitatiivisten astevaihtelusanojen frekvenssin kehitys tuhatta substantiivia kohden. Taitotasolla A1 frekvenssi on 25, A2-tasolla se on 29, taitotasolla B1 se on 34, B2-tasolla frekvenssi on 30, C1-tasolla se on 24 ja viimeisellä tasolla C2 se on 25. Kuten kuviosta 7 näkee, frekvenssi nousee tasaisesti taitotasolta A1 taitotasolle B1, jossa se on korkeimmillaan. Tämän jälkeen se laskee tasaisesti tasoilla B1–C1, jonka jälkeen se pysyy tasaisena viimeiselle taitotasolle C2 siirryttäessä. Tämänkin vaihtelun osalta frekvenssi pysyy suhteellisen tasaisena, koska se vaihtelee vain noin kymmenellä sanalla tuhatta sanaa kohden eri taitotasoilla.

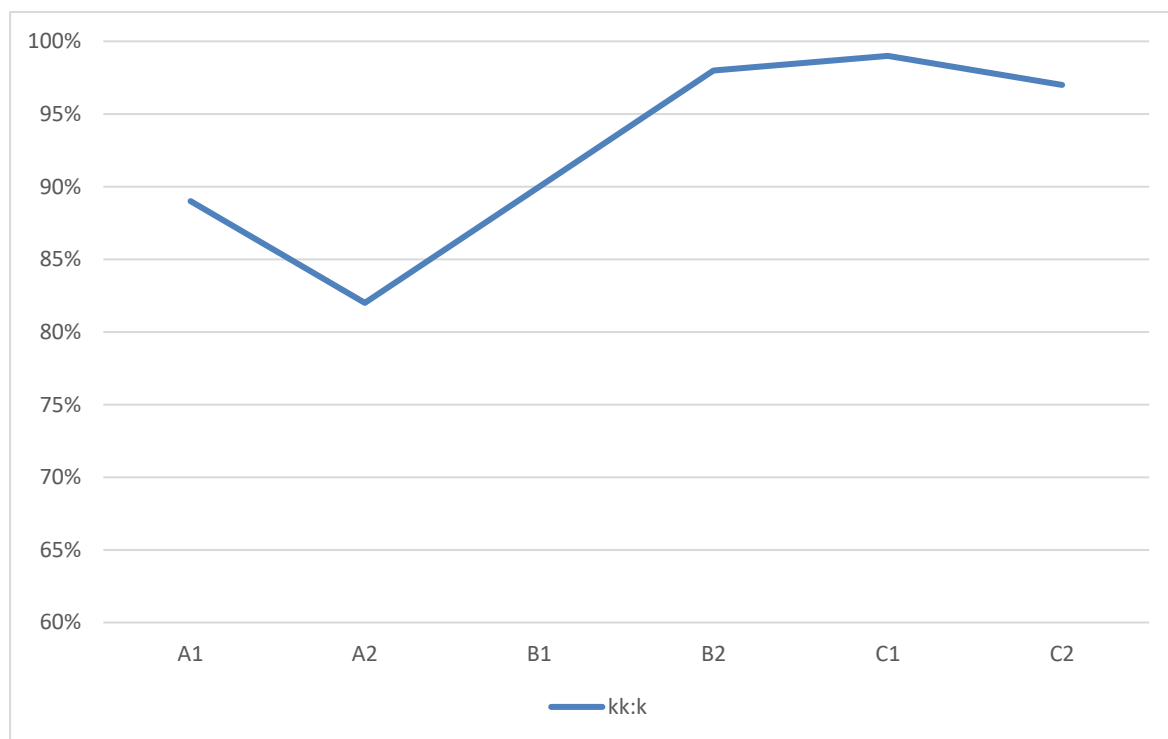
#### 4.2.2 Vaihteluparin jakauma ja tarkkuus

Kvantitatiivisen vaihtelun sanoissa ei luonnollisesti ole kuin yksi vaihtelupari, koska kaikki tämän vaihtelun alaisista sanoista sisältävät *kk* : *k* -vaihtelua. Näin ollen taulukosta 6 näkee siis kyseisen taitotason kokonaisnorminmukaisuuden, koska kaikki sanat sisältävät tätä vaihteluparia.

Taulukko 6. Kvantitatiivisen vaihtelun sanojen vaihteluparin jakauma ja tarkkuus.

	<b>A1</b>	<b>A2</b>	<b>B1</b>	<b>B2</b>	<b>C1</b>	<b>C2</b>		
	n=27	n=100	n=100	n=100	n=100	n=72	Yhteensä vaihteluparia	NM
<b>kk : k</b>	<b>89 %</b> <b>(24)</b>	<b>82 %</b> <b>(82)</b>	<b>90 %</b> <b>(90)</b>	<b>98 %</b> <b>(98)</b>	<b>99 %</b> <b>(99)</b>	<b>97 %</b> <b>(70)</b>	<b>100 %</b> <b>(499)</b>	<b>93 %</b> <b>(463)</b>

Taitotasolla A1 norminmukaisuus on siis 89 %, tasolla A2 se on 82 %, B1-tasolla se on 90 %, B2-tasolla norminmukaisuus on 98 %, C1-tasolla 99 % ja C2-tasolla 97 %. Tästä siis näkee, että millään taitotasolla ei ole osattu muodostaa kaikkia muotoja astevaihtelun valossa oikein. Lähimmäksi 100 % päästään tosin taitotasolla C1, koska tällä tasolla on vain yksi virheellinen muoto. Kaikista otannan kvantitatiivisen vaihtelun alaisista sanoista on osattu muodostaa 93 % oikein.



Kuvio 8. Kvantitatiivisen vaihtelun sanojen vaihteluparin norminmukaisuuden kehitys taitotasoin. Kuten kuviosta 8 näkee, norminmukaisuus ei etene lineaarisesti kaikkien taitotasojen läpi. Norminmukaisuus on lähellä 90 % jo ensimmäisellä tasolla, jonka jälkeen se laskee jyrkästi alas, josta se nousee taas melko tasaisesti ylös. Korkeimmalla se on taitotasolla C1, jonka jälkeen se laskee vielä vähän taitotasolla C2. Korkeampi norminmukaisuus tasolla A1 kuin A2 voi selittyä sillä, että taitotasolla A1 on niin vähän sanoja. Samansuuntaisesta syystä voi johtua myös norminmukaisuuden lasku taitotasolla C2, koska tällä tasolla on vähemmän sanoja kuin taitotasolla C1, jolloin jo muutama virheellinen muoto laskee

norminmukaisuusprosenttia enemmän. Kokonaisuudessaan tämän vaihtelun osalta ollaan siis jo hallinnan puolella, koska millään taitotasolla norminmukaisuusprosentti ei ole alle 80.

#### 4.2.3 Sijamuotojen jakauma ja tarkkuus

Otannassa on tämänkin vaihtelutyypin osalta mukana myös possessiivisuffiksillisia ja liitepartikkeleja sisältäviä muotoja pelkässä sijamuodossa olevien sanojen lisäksi. Tässä vaihtelutyypissä esiintyy sanoja kaikissa muissa sijamuodoissa paitsi vajaakäyttöisissä sijoissa. Kolmen eri sijamuodon kanssa esiintyy myös possessiivisuffiksillisia muotoja, ja kahden sijamuodon kanssa esiintyy myös liitepartikkelia *-kin*. Näistä kaikista on kuitenkin osattu muodostaa norminmukaiset muodot pois lukien yksi genetiivi + possessiivisuffiksi -muoto.

Taulukko 7. Kvantitatiivisen vaihtelun sanojen sijamuotojen jakauma ja tarkkuus.

	<b>A1</b>		<b>A2</b>		<b>B1</b>		<b>B2</b>		<b>C1</b>		<b>C2</b>		<b>Yhteensä sijamuotoa</b>	
	n=27		n=100		n=100		n=100		n=100		n=72		määrä	NM
	määrä	NM	määrä	NM	määrä	NM	määrä	NM	määrä	NM	määrä	NM	määrä	NM
<b>Genetiivi</b>	26 % (7)	<b>86 % (6)</b>	23 % (23)	<b>65 % (15)</b>	15 % (15)	<b>100 % (15)</b>	21 % (21)	<b>95 % (20)</b>	27 % (27)	<b>96 % (26)</b>	37,5 % (27)	<b>96 % (26)</b>	24 % (120)	<b>90 % (108)</b>
<b>Partitiivi</b>	48 % (13)	<b>85 % (11)</b>	35 % (35)	<b>89 % (31)</b>	29 % (29)	<b>90 % (26)</b>	22 % (22)	<b>95 % (21)</b>	16 % (16)	<b>100 % (16)</b>	24 % (17)	<b>100 % (17)</b>	26 % (132)	<b>92 % (122)</b>
<b>Inessiivi</b>	4 % (1)	<b>100 % (1)</b>	13 % (13)	<b>85 % (11)</b>	23 % (23)	<b>91 % (21)</b>	21 % (21)	<b>100 % (21)</b>	16 % (16)	<b>100 % (16)</b>	11 % (8)	<b>100 % (8)</b>	16 % (82)	<b>95 % (78)</b>
<b>Elatiivi</b>	4 % (1)	<b>100 % (1)</b>	3 % (3)	<b>67 % (2)</b>	5 % (5)	<b>100 % (5)</b>	7 % (7)	<b>100 % (7)</b>	11 % (11)	<b>100 % (11)</b>	4 % (3)	<b>100 % (3)</b>	6 % (30)	<b>97 % (29)</b>
<b>Illatiivi</b>	-	-	7 % (7)	<b>71 % (5)</b>	10 % (10)	<b>90 % (9)</b>	11 % (11)	<b>100 % (11)</b>	8 % (8)	<b>100 % (8)</b>	11 % (8)	<b>87,5 % (7)</b>	9 % (44)	<b>91 % (40)</b>
<b>Adessiivi</b>	19 % (5)	<b>100 % (5)</b>	7 % (7)	<b>100 % (7)</b>	6 % (6)	<b>100 % (6)</b>	11 % (11)	<b>100 % (11)</b>	7 % (7)	<b>100 % (7)</b>	5,5 % (4)	<b>100 % (4)</b>	8 % (40)	<b>100 % (40)</b>
<b>Ablatiivi</b>	-	-	-	-	-	-	1 % (1)	<b>100 % (1)</b>	1 % (1)	<b>100 % (1)</b>	-	-	0,4 % (2)	<b>100 % (2)</b>
<b>Allatiivi</b>	-	-	4 % (4)	<b>100 % (4)</b>	4 % (4)	<b>75 % (3)</b>	-	-	3 % (3)	<b>100 % (3)</b>	1 % (1)	<b>100 % (1)</b>	2 % (12)	<b>92 % (11)</b>
<b>Essiivi</b>	-	-	5 % (5)	<b>80 % (4)</b>	4 % (4)	<b>75 % (3)</b>	4 % (4)	<b>100 % (4)</b>	5 % (5)	<b>100 % (5)</b>	1 % (1)	<b>100 % (1)</b>	4 % (19)	<b>89 % (17)</b>
<b>Transla- tiivi</b>	-	-	1 % (1)	<b>100 % (1)</b>	2 % (2)	<b>50 % (1)</b>	1 % (1)	<b>100 % (1)</b>	3 % (3)	<b>100 % (3)</b>	-	-	1 % (7)	<b>86 % (6)</b>
<b>Genetiivi + partikkeli -kin</b>	-	-	1 % (1)	<b>100 % (1)</b>	-	-	-	-	-	-	-	-	0,2 % (1)	<b>100 % (1)</b>
<b>Partitiivi + possessiivi -suffiksi</b>	-	-	1 % (1)	<b>100 % (1)</b>	1 % (1)	<b>100 % (1)</b>	-	-	-	-	3 % (2)	<b>100 % (2)</b>	1 % (4)	<b>100 % (4)</b>
<b>Genetiivi + possessiivi -suffiksi</b>	-	-	-	-	1 % (1)	<b>0 %</b>	-	-	2 % (2)	<b>100 % (2)</b>	1 % (1)	<b>100 % (1)</b>	1 % (4)	<b>75 % (3)</b>
<b>Partitiivi + partikkeli -kin</b>	-	-	-	-	-	-	1 % (1)	<b>100 % (1)</b>	-	-	-	-	0,2 % (1)	<b>100 % (1)</b>
<b>Elatiivi + possessiivi -suffiksi</b>	-	-	-	-	-	-	-	-	1 % (1)	<b>100 % (1)</b>	-	-	0,2 % (1)	<b>100 % (1)</b>
<b>Yhteensä</b>		<b>89 % (24)</b>		<b>82 % (82)</b>		<b>90 % (90)</b>		<b>98 % (98)</b>		<b>99 % (99)</b>		<b>97 % (70)</b>		

Taitotasolla A1 on sanoja vain viidessä eri sijamuodossa. Nämä ovat genetiivi (26 %), partitiivi (48 %), inessiivi (4 %), elatiivi (4 %) ja adessiivi (19 %). Näistä virheellisiä muotoja on vain genetiivissä ja partitiivissa, joiden norminmukaisuusprosentit ovat 86 % ja 85 %. Virheellisiä muotoja on siis kolme, jotka ovat genetiivin osalta *säkin* ja partitiivin osalta *kukata* ja *viikoa*. Muodossa *säkin* ei siis ole käytetty astevaihtelua, koska sanassa on yhä kaksi klusiilia yhden klusiilin sijaan, vaikka näin kuuluisi olla, koska kyseessä on genetiivimuotoinen sana. Molemmissa partitiivimuodoissa on vain yksi *k*, vaikka partitiivissa ei ole astevaihtelua, jolloin näissä sanoissa kuuluisi edelleen olla kaksi klusiilia. Muodossa *kukata* on myös väärä partitiivin päätte, jonka valinta on voinut hämätä oppijaa käyttämään astevaihtelua. Näin ollen tällä tasolla oppijoille haastavin sijamuoto on partitiivi, ja sanoissa on käytetty astevaihtelua, vaikka näin ei norminmukaisesti kuuluisi tehdä. Tällä tasolla on myös muutamia muulla tavalla huomiota herättäneitä muotoja, jotka ovat *villakolla*, *latikon*, *känykkää*, *musiikkiaa* ja *purgist*. Näissä muodoissa on lähinnä mitä luultavimmin kirjoitusvirheitä, mutta muodossa *musiikkiaa* on ikään kuin kaksi *a*:ta partitiivin päätteessä ja *purgist* muodossa elatiivin päätteestä puuttuu *a*, ja *k*:n sijasta muodossa on *g*. Astevaihtelun valossa nämä kaikki muodot ovat kuitenkin norminmukaisia. Tällä tasolla kaikista muodoista oikein on 89 %.

Tasolla A2 jakauma on jo paljon laajempi, ja mukana on ensimmäisen kerran illatiiveja, allatiiveja, essiivejä, translatiiveja ja possessiivisuffiksi sekä liitepartikkeli. Tällä tasolla genetiivejä on 23 %, partitiiveja 35 %, inessiivejä on 13 %, illatiiveja ja adessiiveja on 7 %, essiivejä on 5 %, allatiiveja on 4 %, elatiiveja on 3 %, translatiiveja, genetiivi + *-kin* - ja partitiivi + possessiivisuffiksi -muotoja on kutakin 1 %. Norminmukaisuusprosentit vaihtelevat myös paljon. Kaikilta muodoiltaan norminmukaisia ovat sanat adessiivissa, allatiivissa, translatiivissa ja genetiivi + *-kin* - ja partitiivi + possessiivisuffiksi -muodoissa. Partitiiveista norminmukaisia on 89 %, inessiiveistä 85 %, essiiveistä 80 %, illatiiveista 71 %, elatiiveista 67 % ja genetiiveistä 65 %. Tällä tasolla on yhteensä 18 norminvastaista sanamuotoa. Ne ovat partitiiveista *viikoa*, *viikoa*, *musiikita* ja *karkeä*. Näissä kaikissa muodoissa on käytetty astevaihtelua, vaikka näin ei olisi kuulunut tehdä. Muoto *musiikita* on virheellinen senkin takia, että siinä on valittu väärä partitiivin päätte, joka on voinut saada oppijan muodostamaan norminvastaisen muodon. Muoto *karkeä* olisi norminmukaisesta karkkia, jonka voi päätellä sanan kontekstista, jolloin siinäkin on muitakin norminvastaisuutta aiheuttavia tekijöitä. Inessiivimuotoiset norminvastaiset muodot ovat *kirkkossa* ja *kirkkossa*. Nämä ovat siis samanmuotoiset, ja niissä ei ole käytetty astevaihtelua, vaikka inessiivin päätte

edellyttäisi sitä. Essiivimuotoisten sanojen virheellinen muoto on *viikona*, eli tässä muodossa on käytetty astevaihtelua, vaikka essiivissä ei sitä ole.

Illatiivisanoista virheelliset muodot ovat *kirkoon* ja *diskotekiin*. Näissä molemmissa on käytetty astevaihtelua, vaikka ei olisi kuulunut. Jälkimmäisesti muodosta puuttuu myös toinen *e*, ja sana on vanhahtava, jolloin oppija ei välttämättä ole törmännyt edes sanan perusmuotoon useasti. Elatiivisanojen norminvastainen muoto on *musiikkista*, jossa ei ole käytetty astevaihtelua, vaikka olisi pitänyt. Genetiivisanojen norminvastaiset muodot ovat *Pekkan*, *Pekkan*, *Pekkan*, *Jukkan*, *Jukkan*, *musiikkien*, *musiikkien* ja *pankkien*. Näistä muodoista viisi on etunimiä, jonka takia niitä ei välttämättä ole taivutettu niin, että niissä on astevaihtelua. Kolme jälkimmäistä ovat yleissanoja, mutta niissä ei ole käytetty astevaihtelua, vaikka genetiivin päätte sitä edellyttäisi. Muodossa *musiikkien* on myös ikään kuin kaksi sijamuodon päätettä peräkkäin tai sitten vain kirjoitusvirhe ylimääräisen *a:n* muodossa. Tällä koko taitotasolla norminmukaisuusprosentti on 82 %. Tällä tasolla haastavin sijamuoto on genetiivi, mutta silti kaikista virheellisistä muodoista 11 on sellaisia, joissa ei ole käytetty astevaihtelua, vaikka olisi kuulunut. Näin ollen jopa seitsemän muotoa on sellaisia, joissa ei olisi kuulunut käyttää astevaihtelua, mutta silti sitä on käytetty. Tällainen jakauma on poikkeuksellinen verrattuna kvalitatiivisen vaihtelun alaisiin sanoihin.

Taitotason B1 jakauma on hyvin samanlainen kuin aiemmallakin taitotasolla, tosin tällä tasolla ei ole liitepartikkeleja, mutta mukana on possessiivisuffiksilliset muodot sekä genetiivistä että partitiivista. Tällä tasolla on myös enemmän inessiivejä kuin genetiivejä, jolla se eroaa edellisestä tasosta. Tällä tasolla on eniten partitiiveja (29 %), toiseksi eniten inessiivejä (23 %), sitten genetiivejä (15 %) ja sitten illatiiveja (10 %). Tämän jälkeen eniten on adessiiveja (6 %), sitten elatiiveja (5 %), allatiiveja ja essiivejä on molempia 4 % ja 2 % on translatiiveja. Tämän lisäksi otannassa on possessiivisuffiksilliset muodot genetiivistä ja partitiivista, joita kumpiakin on 1 %.

Tällä tasolla kaikki muodot ovat oikein genetiivissä, elatiivissa, adessiivissa ja partitiivi + possessiivisuffiksi -muodossa. Inessiivimuotoisista sanoissa oikein on 91 %, ja virheelliset muodot ovat *Kreikkassa* ja *kreikkassa*. Näissä kummassakaan ei siis ole käytetty astevaihtelua, vaikka inessiivin päätte sitä edellyttäisi. Partitiivisanoista oikein on 90 %, ja nämä virheelliset muodot ovat *kukata*, *kukatta* ja *muusiikia*. Molemmissa *kukka*-sanan partitiivimuodossa on väärä partitiivin päätte, joka on voinut vaikuttaa sanan taivutukseen. Jälkimmäisessä päätte muistuttaa abessiivin päätettä, mutta muodolla on mitä

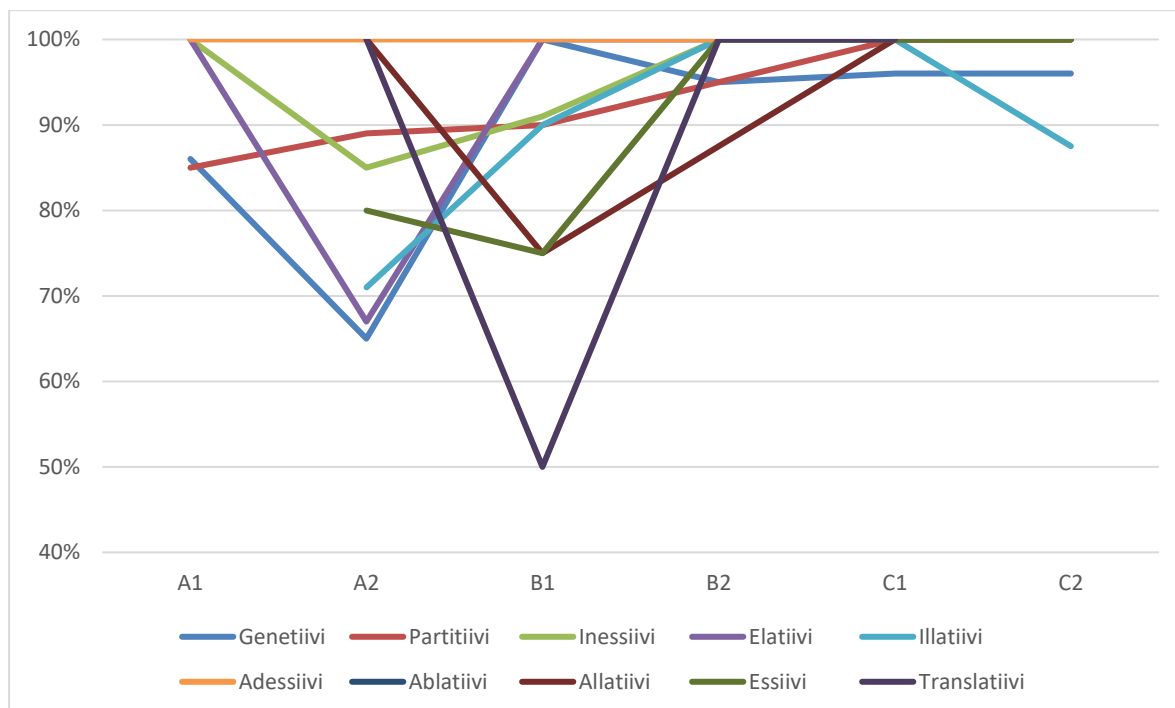
todennäköisimmin haettu partitiivimuotoa, kun sanan konteksti otetaan huomioon. Muodossa *muusiikia* on astevaihteluvirheen lisäksi myös yksi ylimääräinen *u*, joka voi olla kirjoitusvirhe. Illatiivisanoja on oikein 90 % ja norminvastainen muoto on *paikaan*. Tässä on käytetty astevaihtelua, vaikka ei olisi kuulunut. Allatiivimuotoisista sanoista oikein on 75 %, ja norminvastainen muoto on *tupakkalle*, jossa ei ole käytetty astevaihtelua, vaikka norminmukaisesti niin olisi kuulunut tehdä. Myös essiivimuotoisista sanoista oikein on 75 %, ja norminvastainen muoto on *keskiviikona*, jossa on käytetty astevaihtelua, vaikka ei olisi kuulunut. Translatiivisanoista oikein on 50 %, ja norminvastainen muoto on *kukkoksi*. Tässäkään ei ole käytetty astevaihtelua, vaikka olisi kuulunut. Genetiivi + possessiivisuffiksi -muotoja on vain yksi, joka on norminvastainen, joten norminmukaisuusprosentti on 0. Tämä muoto on *serkuuni*, jossa astevaihtelua on käytetty norminvastaisesti, koska possessiivisuffiksien kanssa ei käytetä astevaihtelua. Tällä tasolla on myös kaksi muuta huomiota herättänyttä muotoa, jotka ovat *tutikin* ja *vikkona*, jotka ovat todennäköisesti kirjoitusvirheellisiä muotoja, mutta astevaihtelua näissä on käytetty oikein. Tällä tasolla norminmukaisuusprosentti on 90. Oppijoille haastavin sijamuoto tällä tasolla prosentuaalisesti on translatiivi, mutta määrällisesti eniten virheellisiä muotoja on partitiivissa. Tälläkin tasolla on enemmän virheitä niin päin, että astevaihtelua on käytetty, kun sitä ei kuuluisi käyttää, kuin niin päin, että astevaihtelua ei ole käytetty, vaikka olisi pitänyt.

Seuraavalla taitotasolla B2 jakauma on hieman tasaisempi kuin aiemmillä tasoilla. Tällä tasolla genetiivejä sanoista on 21 %, partitiivejä 22 % ja inessiivejä 21 %. Tämän jälkeen sekä illatiiveja että adessiiveja on 11 %. Elatiiveja on 7 %, essiivejä on 4 % ja ablatiiveja, translatiiveja ja partitiivi + liitepartikkeli -kin -muotoja on kutakin 1 %. Tällä tasolla esiintyy siis ensimmäisen kerran ablatiiveja, mutta jo kahdesti esiintyneitä allatiivimuotoja ei ole ollenkaan. Tällä tasolla kaikki muodot ovat muutoin oikein, mutta sekä genetiivien että partitiivien norminmukaisuus on 95 %. Nämä virheelliset muodot ovat *särkkän* ja *luokaa*. Ensimmäisessä muodossa ei ole käytetty genetiivissä astevaihtelua, ja toisessa on tehty astevaihtelullinen muoto partitiivissa. Mukana oli myös muutamia huomiota herättäneitä muotoja, joissa on lähinnä kirjoitusvirheitä vokaaliharmonian suhteen, ja joista puuttuu muita kirjaimia, mutta astevaihtelun osalta ne olivat oikein. Kokonaisnorminmukaisuus on tällä tasolla 98 %, ja mukana on yksi astevaihteluton muoto ja yksi muoto, jossa sitä on käytetty tarpeettomasti. Näin ollen ei voi sanoa kumpi virhetyyppi on tällä tasolla yleisempi.

Taitotasolla C1 jakauma ei ole enää niin tasainen, ja genetiivimuotoisia sanoja on selkeästi eniten (27 %). Sekä partitiiveja että inessiivejä on 16 %, elatiiveja 11 %, illatiiveja 8 %,

adessiiveja 7 % ja essiivejä 5 %. Näiden lisäksi sekä allatiiveja että translatiiveja on 3 %, genetiivin possessiivisuffiksillisia muotoja on 2 % ja sekä ablatiiveja että elatiivin possessiivisuffiksillisia muotoja on kumpiakin 1 %. Norminmukaisuus on muutoin 100 %, mutta genetiivimuotoisten sanojen norminmukaisuusprosentti on 96 %. Tämän tason virheellinen muoto on *musiikkin*, jossa ei ole käytetty astevaihtelua genetiivissä. Tällä tasolla oli myös muita virheellisiä muotoja, joissa on kirjoitusvirheitä, mutta astevaihtelun osalta ne olivat oikein, jonka takia ne on luokiteltu norminmukaisiksi. Tämän tason kokonaisnorminmukaisuusprosentti onkin 99 %, joka on korkein tämän vaihtelun osalta. Tällä tasolla siis haastavin sijamuoto on genetiivi, ja virhe on tehty niin päin, että astevaihtelua ei ole käytetty genetiivimuotoisessa sanassa.

Korkeimmalla taitotasolla eli tasolla C2 jakauma on taas vähemmän tasaisempi kuin aiemmillä taitotasoilla. Tällä tasolla genetiivejä on 37,5 %, partitiiveja 24 % ja inessiivejä sekä illatiiveja on 11 %. Näiden lisäksi adessiiveja on 5,5 %, elatiiveja on 4 % ja partitiivin possessiivisuffiksillisia muotoja on 3 %. Näiden lisäksi otannassa on tällä tasolla 1 % allatiiveja, essiivejä ja genetiivin possessiivisuffiksillisia muotoja. Ablatiiveja tai translatiiveja ei siis ole laisinkaan korkeimmalla taitotasolla. Norminmukaisuudet ovat tällä tasolla muutoin 100 %, mutta virheellisiä muotoja on sekä genetiiveissä (96 % oikein) että illatiiveissa (87,5 % oikein). Nämä virheelliset muodot ovat *nurkaan* ja *kirkkon*. Ensimmäisessä muodossa on muodostettu astevaihtelullinen muoto illatiivissa, ja toisessa on muodostettu genetiivi ilman astevaihtelua. Tälläkin tasolla oli muutamia kirjoitusvirheellisiä muotoja näiden lisäksi, mutta niissä astevaihtelua on käytetty oikein. Tällä tasolla kokonaisnorminmukaisuusprosentti on 97 %. Tällä tasolla haastavimmat sijamuodot ovat siis genetiivi ja illatiivi, ja mukana on sekä astevaihtelun käyttöä että käyttämättömyyttä väärissä sijamuodoissa.



Kuvio 9. Kvantitatiivisen vaihtelun sanojen sijamuotojen norminmukaisuuden kehitys taitotasoin. Kuten kuvio 9 näkee, sijamuotojen norminmukaisuudet vaihtelevat melko paljon taitotasolta toiselle siirryttäessä. Kokonaisuudessaan sekä adessiivin että ablatiivin osalta oppijat ovat muodostaneet kaikki muodot oikein, jolloin tämä näkyy kuviossa 9 suorana viivana yläreunassa. Genetiivin norminmukaisuus lähtee noin 85 %:sta, jonka jälkeen se laskee radikaalisti, josta se nousee jyrkästi 100 %:iin ja sen jälkeen laskee hieman, jossa se pysyy loppuilla taitotasoin. Partitiivin norminmukaisuus nousee hyvin tasaisesti alimmilta tasoilta siirryttäessä ylemmille tasoille ja saavuttaa 100 %:in tasolla C1, jossa se pysyy vielä viimeiselläkin tasolla. Inessiivi lähtee 100 %:sta, jonka jälkeen se laskee reippaasti seuraavalla tasolla, ja nousee taas tasaisesti saavuttaen 100 %:n taitotasolla B2. Allatiivin norminmukaisuus lähtee samoin 100 %:sta, tosin vasta tasolta A2. Tämän jälkeen se kuitenkin laskee merkittävästi 75 %:n tasolle, jonka jälkeen se nousee tasaisesti ja on taas 100 % tasolla C1, jossa se pysyy myös tasolla C2. Elatiivin norminmukaisuus pysyy muutoin koko ajan 100 %:ssa, mutta se tekee merkittävän notkahduksen taitotasolla A2. Essiivin norminmukaisuus lähtee 80 %:n tasolta ja vasta taitotasolta A2. Tämän jälkeen se laskee entisestään, jonka jälkeen se nousee suoraan 100 %:iin, jossa se pysyy. Illatiivin norminmukaisuus lähtee noin 70 %:sta, josta se nousee tasaisesti 100 %:iin taitotasolla B2, jossa se on myös tasolla C1, jonka jälkeen se kuitenkin laskee vielä viimeisellä tasolla 90 %:n tienoille. Translatiivia ei esiinny ollenkaan ensimmäisellä eikä viimeisellä tasolla. Se kuitenkin lähtee taitotasolla

A2 100 %:sta, jonka jälkeen se tippuu hyvin radikaalisti 50 %:iin, jonka jälkeen se on taas loppuilla taitotasoilla 100 %.

Sijamuotokohtaiset kokonaisnorminmukaisuusprosentit vaihtelevat melko paljon.

Kokonaisuudessaan jokainen muoto on oikein sanoissa, jotka ovat ablatiivissa, adessiivissa, partitiivi – ja elatiivi + possessiivisuffiksi -muodoissa sekä *-kin* partikkelin kanssa olevissa partitiivi- ja genetiivimuodoissa. Kaikista elatiiveista on oikein 97 %, inessiiveistä 95 %, partitiiveista ja allatiiveista 92 %, illatiiveista 91 %, genetiiveistä 90 %, essiiveistä 89 %, translatiiveista 86 % ja genetiivi + possessiivisuffiksi -muodoista 75 %. Jos ajatellaan 80 % hallintaa osoittavaa rajaa, tämän alittaa vain genetiivi + possessiivisuffiksi -muoto, tosin näitä on koko otannassa tämän vaihtelun osalta vain kolme. Muutoin tämän rajan alittaa muutaman sijamuodon sanat taitotasoilla A2 ja B1.

#### 4.2.4 Sanojen pituuden jakauma ja tarkkuus

Kvantitatiivisen vaihtelun alaiset sanat ovat pidempiä kuin kvalitatiivisen vaihtelun alaiset sanat. Tämä näkyy yksinkertaisesti jo siinä, että tämän vaihtelun sanoissa on paljon enemmän viisitavuisia muotoja, ja mukana on myös muutama kuusitavuinen sana. Eniten on kuitenkin kolmetavuisia (49 %), sitten kaksitavuisia (24 %), sitten neljätavuisia (19 %), sitten viisitavuisia (6 %) ja viimeiseksi kuusitavuisia sanoja (0,4 %). Molemmat kuusitavuiset sanat ovat kuitenkin oikein, viisitavuisista oikein on 97 %, sekä kolme- että neljätavuisista sanoista oikein on 94 % ja kaksitavuisista sanoista on oikein 89 %. Näin ollen tämän vaihtelun sanoista haastavimpia astevaihtelun kannalta oppijoille on kaksitavuiset sanat.

Taulukko 8. Kvantitatiivisen vaihtelun sanojen pituuden jakauma ja tarkkuus.

	<b>A1</b>		<b>A2</b>		<b>B1</b>		<b>B2</b>		<b>C1</b>		<b>C2</b>		Yhteensä pituuksia	
	n=27		n=100		n=100		n=100		n=100		n=72		Mää- rä	NM
	määrä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM
<b>2 tavua</b>	30 % (8)	<b>87,5</b> % (7)	23 % (23)	<b>70</b> % (16)	17 % (17)	<b>94</b> % (16)	21 % (21)	<b>90</b> % (19)	29 % (29)	<b>100</b> % (29)	33 % (24)	<b>92</b> % (22)	24 % (122 )	<b>89 %</b> (109)
<b>3 tavua</b>	59 % (16)	<b>87,5</b> % (14)	41 % (41)	<b>85</b> % (34)	48 % (48)	<b>87,5</b> % (42)	55 % (55)	<b>100</b> % (55)	47 % (47)	<b>98 %</b> (46)	56 % (40)	<b>100</b> % (40)	49 % (247 )	<b>94 %</b> (231)
<b>4 tavua</b>	11 % (3)	<b>100</b> % (3)	31 % (31)	<b>87</b> % (27)	24 % (24)	<b>92</b> % (22)	17 % (17)	<b>100</b> % (17)	16 % (16)	<b>100</b> % (16)	8 % (6)	<b>100</b> % (6)	19 % (97)	<b>94 %</b> (91)
<b>5 tavua</b>	-	-	5 % (5)	<b>100</b> % (5)	10 % (10)	<b>90</b> % (9)	7 % (7)	<b>100</b> % (7)	7 % (7)	<b>100</b> % (7)	3 % (2)	<b>100</b> % (2)	6 % (31)	<b>97 %</b> (30)
<b>6 tavua</b>	-	-	-	-	1 % (1)	<b>100</b> % (1)	-	-	1 % (1)	<b>100</b> % (1)	-	-	0,4 % (2)	<b>100</b> % (2)
<b>Yh- teen- sä</b>		<b>89</b> % (24)		<b>82</b> % (82)		<b>90</b> % (90)		<b>98</b> % (98)		<b>99 %</b> (99)		<b>97</b> % (70)		

Kuten taulukosta 8 näkee, taitotasolla A1 on sekä kaksi-, kolme- että neljätavuisia sanoja.

Eniten on kuitenkin kolmetavuisia (59 %), sitten kaksitavuisia (30 %) ja viimeiseksi neljätavuisia (11 %) sanoja. Ensimmäisellä tasolla on siis lyhyimmät sanat tämän vaihtelutyypin osalta. Kaikki neljätavuiset sanat ovat oikein, ja sekä kaksi- että kolmetavuisten sanojen norminmukaisuusprosentti on 87,5. Kaksitavuisten sanojen virheellinen muoto on *säkin*, ja kolmetavuisten sanojen virheelliset muodot ovat *viikoa* ja *kukata*. Kaksitavuinen sana on genetiivissä, jolloin siihen lisätään yksi konsonantti, jolloin viimeinen tavu on muotoa CVC. Molemmat kolmetavuiset sanat ovat partitiiveja, jolloin näihin tulisi lisätä vain yksi vokaali, jolloin sanaan ei tule umpitavua, joka laukaisisi astevaihtelun. Muotoon *kukata* on kuitenkin lisätty väärä partitiivin päätte, jolloin sanaan on lisätty päätte, jossa on konsonantti ja vokaali. Tällaisen päätteen ei kuitenkaan astevaihtelun pääsäännön mukaan tulisi laukaista astevaihtelua. Tällä tasolla yhtä vaikeita ovat siis kaksi- ja kolmetavuiset sanat, mutta määrällisesti kolmetavuisissa sanoissa on enemmän virheitä. Tällä tasolla vaikeimpia ovat sellaiset sanat, joihin kuuluisi lisätä yksi vokaali, joka muuttaa tavun muodosta CV muotoon CVV.

Taitotasolla A2 on ensimmäisen kerran viisitavuisia sanoja. Eniten on kuitenkin kolmetavuisia sanoja (41 %), sitten neljätavuisia sanoja (87 %), sitten kaksitavuisia (23 %) ja viimeiseksi viisitavuisia (5 %) sanoja. Kaikki viisitavuiset sanat ovat oikein, neljätavuisista

sanoista oikein on 87 %, kolmetavuisista sanoista 85 % ja kaksitavuisista vain 70 %. Näin ollen kaksitavuisten sanojen norminmukaisuus ei vielä ylitä 80 % -hallintarajaa.

Neljätavuisten sanojen virheelliset muodot ovat *musiikita*, *musiikkian*, *musiikkista* ja *diskotekiin*. Muutoin näiden kaikkien sanojen tavumäärä kasvaa päätteen lisäämisen jälkeen yhdellä tavulla, mutta muodon *diskotekiin* tavumäärä pysyy samana. Muoto *musiikita* on siis partitiivi, jolloin oikea muoto olisi *musiikkia*, eli sanaan on lisätty norminvastainen partitiivin päätte, joka myös muuttaa tavurakennetta. Sana ei kuitenkaan muutu sellaiseksi, että sanassa kuuluisi siltikään olla astevaihtelua. Muoto *musiikkian* on genetiivi, mutta tähänkin sanaan on lisätty norminvastainen päätte, jolloin oikean yhden konsonantin lisäämisen sijaan sanaan on lisätty vokaali ja konsonantti. Tällöin sanan tavurakenne muuttuu, eikä sanassa tällä päätteellä kuuluisikaan olla astevaihtelua. Muodossa *musiikkista* on norminmukainen päätte, joka on muotoa CCV. Tällöin sanassa kuuluisi olla astevaihtelua, koska viimeinen tavu muuttuu muodosta CV muotoon CVC ja toiseksi viimeiseksi tavuksi. Muoto *diskotekiin* on illatiivi, jolloin päätte on muotoa CV. Illatiivisanoissa ei ole astevaihtelua, joka tosin ei johdu tavurakenteesta.

Kolmetavuisten sanojen virheelliset muodot ovat *viikoa*, *viikoa*, *musiikkin*, *viikona*, *kirkkossa*, *kirkkossa* ja *karkeä*. Muihin sanoihin tulee yksi tavu lisää, mutta muodon *musiikkin* tavumäärä pysyy samana. Muodot *viikoa*, *viikoa* ja *karkeä* ovat partitiiveja, jolloin tavu CV muuttuu muotoon CVV, eikä näissä sanoissa näin ollen tulisi olla astevaihtelua. Muotoon *musiikkin*, joka on genetiivi, lisätään yksi konsonantti, jolloin tavurakenne muuttuu muodosta CV muotoon CVC, jolloin tässä sanassa kuuluisi olla astevaihtelua. Muoto *viikona* on essiivi, jolloin sanaan tulee kokonaan uusi tavu lisää, joka on muodoltaan CV. Tämä ei kuitenkaan muuta perusmuodon viimeistä tavua, jolloin sanassa ei kuuluisi olla astevaihtelua. Molemmat muodot *kirkkossa* ovat inessiivejä, jolloin sanaan lisätään päätte, joka on muotoa CCV, ja sanassa pitäisi olla astevaihtelua. Tällöin sanan viimeinen tavu CV muuttuukin sanan keskimmäiseksi tavuksi ja muotoon CVC.

Kaksitavuisten sanojen virheelliset muodot ovat *Pekkan*, *Pekkan*, *Pekkan*, *Jukkan*, *Jukkan*, *kirkoon* ja *pankkin*. Kaikkien näiden muotojen tavumäärät pysyvät samoina. Muodot *Pekkan*, *Jukkan* ja *pankkin* ovat genetiivejä, jolloin sanaan lisätään yksi konsonantti. Tällöin sanojen viimeiset tavut muuttuvat muodosta CV muotoon CVC, jolloin sanoissa kuuluisi olla astevaihtelua. Muoto *kirkoon* on illatiivi, jolloin sanaan lisätään vokaali ja konsonantti. Tällä tasolla haastavimpia sanoja ovat sellaiset sanat, joissa on kaksi tavua, ja joihin lisätään yksi konsonantti, jolloin viimeinen tavu muuttuu muodosta CV muotoon CVC.

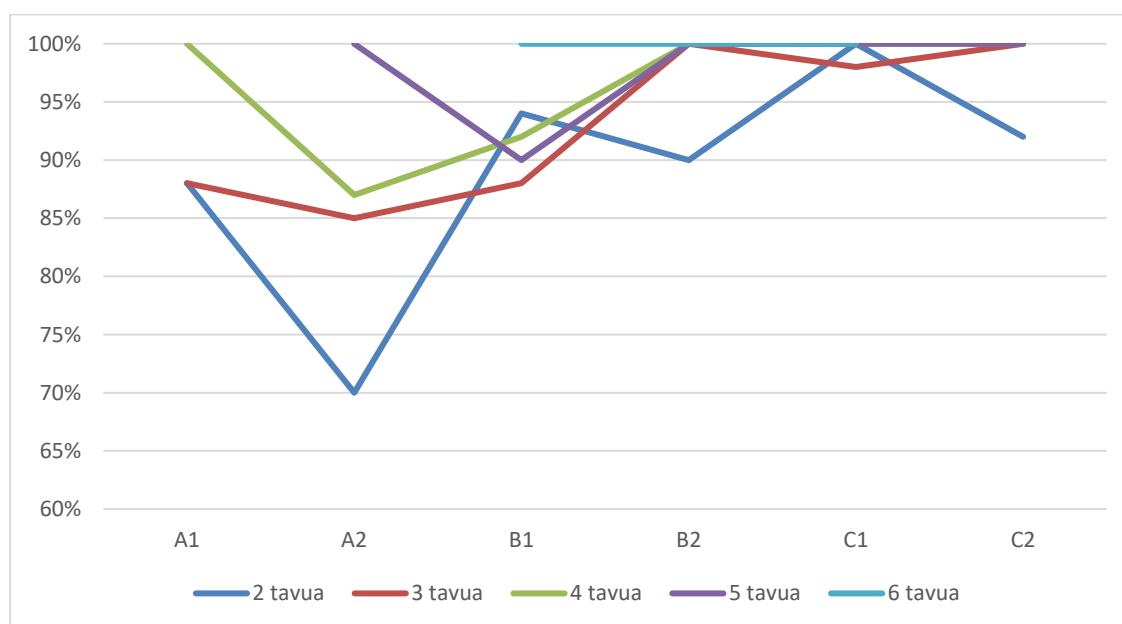
Taitotasolla B1 on mukana myös kuusitavuinen sana ensimmäistä kertaa. Muutoin tämän tason sanoista 48 % on kolmitavuisia, 24 % neljätavuisia, 17 % kaksitavuisia, 10 % viisitavuisia ja kuusitavuisia sanoista on 1 %. Kuusitavuinen muoto on oikein, viisitavuisista oikein on 90 %, neljätavuisista 92 %, kolmetavuisista 87,5 %, ja kaksitavuisista muodoista oikein on 94 %. Kaksitavuisien sanojen virheellinen muoto on *paikaan*, joka on illatiivi, jolloin sanaan lisätään vokaali ja konsonantti, mutta tavumäärä pysyy samana. Sanan viimeinen tavu muuttuu muodosta CV muotoon CVVC, eikä sanassa kuuluisi olla astevaihtelua. Kolmetavuisien sanojen virheelliset muodot ovat *kukata*, *serkuuni*, *kukatta*, *Kreikkassa*, *kreikkassa* ja *kukkoksi*. Näihin kaikkiin sanoihin tulee yksi tavu lisää. Muodot *kukata* ja *kukatta* ovat molemmat partitiiveja, joissa on väärä partitiivin päätte, koska päätteeseen kuuluisi olla vain yksi vokaali. Ensimmäisessä muodossa ei tällaisellakaan päätteellä kuuluisi olla astevaihtelua, mutta jälkimmäisen muodon päätte kyllä laukaisee astevaihtelun, joten tietyllä tavalla muoto ei ole virheellinen. Muoto *serkuuni* on siis genetiivi, jossa on possessiivisuffiksi (CV), eikä siinä kuuluisi olla astevaihtelua. Muotoihin *kreikkassa* ja *kukkoksi* lisätään päätte, joka on muotoa CCV, jolloin se muuttaa viimeisen tavun muodosta CV muotoon CVC ja keskeemmälle sanaan, jolloin sanassa kuuluisi olla astevaihtelua.

Neljätavuisien sanojen virheelliset muodot ovat *tupakkalle* ja *muusiikia*, ja näihin tulee yksi tavu lisää. Ensimmäinen muoto on adessiivi, ja sen päätte on CCV, jolloin se muuttaa sanan viimeisen tavun muodosta CV muotoon CVC ja toiseksi viimeiseksi tavuksi, jolloin sanassa on astevaihtelua. Toinen muoto on partitiivi, jolloin sanaan lisätään vain yksi vokaali ja viimeinen tavu muuttuu muodosta CV muotoon CVV, eikä sanassa kuuluisi olla astevaihtelua. Viisitavuisien sanojen virheellinen muoto on *keskiviikona*, jolloin sanaan on tullut yksi tavu lisää. Muoto on essiivi, jolloin päätte on CV, eikä se vaikuta edeltäviin tavuihin, jolloin sanassa ei kuuluisi olla astevaihtelua. Tällä tasolla haastavimpia sanoja ovat kolmetavuiset sanat, joiden viimeinen tavu CV muuttuu sanan keskimmäiseksi tavuksi ja muotoon CVC.

Taitotasolla B2 ei ole kuusitavuisia sanoja laisinkaan. Muutoin eniten on kolmetavuisia sanoja (55 %), sitten kaksitavuisia (21 %), sitten neljätavuisia (17 %) ja viimeiseksi viisitavuisia (7 %) sanoja. Muutoin kaikki muodot ovat oikein, mutta kaksitavuisissa sanoissa on kaksi virheellistä muotoa, jolloin norminmukaisuusprosentti on 90. Nämä muodot ovat *särkkän* ja *luokaa*. Kumpaankaan muotoon ei tule lisää tavuja. Ensimmäinen sana on genetiivi, jolloin viimeinen tavu muuttuu muodosta CV muotoon CVC, jolloin sanassa kuuluisi olla astevaihtelua. Toinen muoto on partitiivi, jolloin viimeinen tavu muuttuu muodosta CV

muotoon CVV, eikä sanassa kuuluisi olla astevaihtelua. Näin ollen tällä tasolla haastavimpia ovat kaksitavuiset sanat, ja sekä viimeisen tavun muuttuminen umpitavuksi että sen jääminen avotavuksi on yhtä haastavaa, koska molempia virhetyyppejä oli yksi.

Taitotasolla C1 on taas myös kuusitavuisen sana. Muutoin jakauma on seuraavanlainen: kaksitavuisia sanoja on 29 %, kolmetavuisia sanoja on 47 %, neljätavuisia sanoja on 16 %, viisitavuisia sanoja on 7 % ja kuusitavuisia sanoja 1 %. Muutoin kaikki muodot ovat oikein, mutta mukana on yksi virheellinen kolmetavuisen muoto, jolloin norminmukaisuusprosentti on 98 %. Tämä muoto on *musiikkin*, johon ei tule lisää tavuja. Kyseessä on genetiivi, jolloin viimeinen tavu muuttuu muodosta CV muotoon CVC, jolloin sanassa pitäisi olla astevaihtelua. Näin ollen tämä on myös tämän taitotason haastavin tyyppi. Taitotasolla C2 ei ole kuusitavuisia sanoja. Muutoin kaksitavuisia sanoja on 33 %, kolmetavuisia 56 %, neljätavuisia 8 % ja viisitavuisia 3 %. Kaikki muut muodot ovat norminmukaisia, mutta kaksitavuisissa sanoissa on kaksi norminvastaista muotoa, jolloin norminmukaisuusprosentti on 92 %. Nämä muodot ovat *nurkaan* ja *kirkon*, eikä kumpaakaan sanaan tule lisää tavuja. Ensimmäisessä muodossa viimeinen tavu muuttuu muodosta CV muotoon CVVC, eikä tässä kuuluisi olla astevaihtelua. Toisessa muodossa viimeinen tavu muuttuu muodosta CV muotoon CVC, ja tässä kuuluisi olla astevaihtelua. Näin ollen tällä tasolla haastavin sanan pituus on kaksi tavua, ja sekä viimeisen tavun muuttuminen muodosta CV muotoon CVVC ja CVC ovat yhtä haastavia.



Kuvio 10. Kvantitatiivisen vaihtelun sanojen norminmukaisuuden kehitys tavujen määrän mukaan taitotasoittain.

Kuten kuvioista 10 näkee, kolme-, neljä- ja viisitavuisten sanojen norminmukaisuuden kehitys on melko samanlaista. Kolmetavuisten sanojen norminmukaisuus lähtee läheltä 90 %, jonka jälkeen se laskee vähän, josta se nousee aika tasaisesti, ja saavuttaa 100 %:n taitotasolla B2. Tämän jälkeen se tekee vielä pienen notkahduksen, mutta nousee jälleen 100 %:iin. Neljätavuisten sanojen norminmukaisuus lähtee 100 %:sta, jonka jälkeen se putoaa reippaasti, mutta kehittyy taas tasaisesti, kunnes se on 100 % tasolla B2. Viisitavuisten sanojen norminmukaisuus lähtee 100 %:sta, mutta vasta tasolta A2. Tämän jälkeen se tippuu reippaasti, mutta nousee taas seuraavalla tasolla 100 %:iin. Kuusitavuiset sanat ovat koko esiintymisensä ajan 100 %:ssa. Kaksitavuisten sanojen norminmukaisuus on hyvin vaihteleva. Se lähtee suhteellisen korkealta läheltä 90 %, jonka jälkeen se tippuu radikaalisti. Tästä se nousee taas reippaasti, jonka jälkeen se taas laskee. Tämän jälkeen tasolla C1 norminmukaisuus on täydet 100 %, mutta senkin jälkeen norminmukaisuus tippuu jälleen lähelle 90 %.

#### 4.2.5 Sanatyyppien jakauma ja tarkkuus

Kvantitatiivisen vaihtelun sanojen jakauma eri sanatyyppeihin on hieman erilainen kuin kvalitatiivisissa sanoissa, koska tämän vaihtelun alaisissa sanoissa ei kaikilla taitotasolla ole kaikkien sanatyyppien edustajia. Kolmella ensimmäisellä taitotasolla ei ole lainkaan LEHTI-tyyppin sanoja. Tämän vaihtelun sanoista 70 % on KATTO-tyyppin sanoja, 26 % PANKKI-tyyppin sanoja ja vain 3 % on LEHTI-tyyppin sanoja. Pienin norminmukaisuusprosentti on 92 %, joka on PANKKI-sanoissa. Sen jälkeen on KATTO-sanat, joiden norminmukaisuusprosentti on 93 %. Kaikki LEHTI-tyyppin sanat ovat oikein.

Taulukko 9. Kvantitatiivisen vaihtelun sanojen sanatyypin jakauma ja tarkkuus.

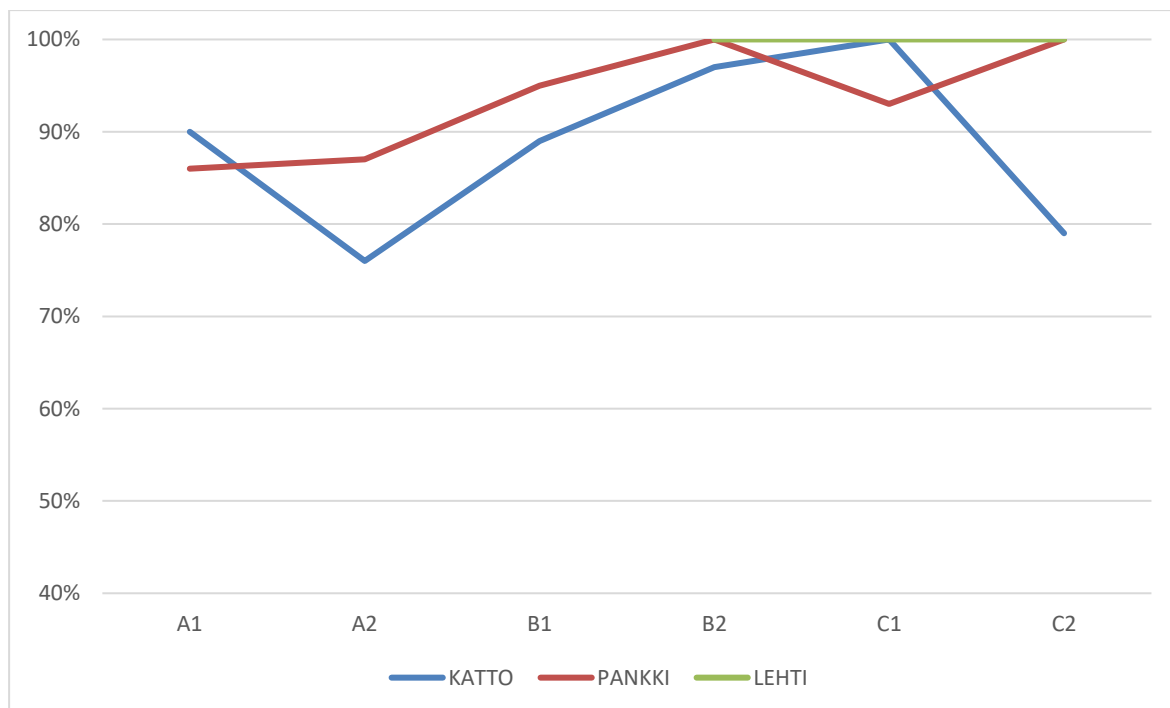
	A1		A2		B1		B2		C1		C2		Yhteensä sanatyyppejä	
	n=27		n=100		n=100		n=100		n=100		n=72		määrä	NM
	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM	Mää- rä	NM		
<b>KATTO</b>	74 % (20)	<b>90</b> % (18)	46 % (46)	<b>76 %</b> (35)	80 % (80)	<b>89</b> % (71)	65 % (65)	<b>97</b> % (63)	81 % (81)	<b>100</b> % (81)	82 % (59)	<b>79</b> % (57)	70 % (351)	<b>93</b> % (325)
<b>PANKKI</b>	26 % (7)	<b>86</b> % (6)	54 % (54)	<b>87 %</b> (47)	20 % (20)	<b>95</b> % (19)	30 % (30)	<b>100</b> % (30)	15 % (15)	<b>93 %</b> (14)	8 % (6)	<b>100</b> % (6)	26 % (132)	<b>92</b> % (122)
<b>LEHTI</b>	-	-	-	-	-	-	5 % (5)	<b>100</b> % (5)	4 % (4)	<b>100</b> % (4)	10 % (7)	<b>100</b> % (7)	3 % (16)	<b>100</b> % (16)
<b>Yhteensä</b>		<b>89</b> % (24)		<b>82 %</b> (82)		<b>90</b> % (90)		<b>98</b> % (98)		<b>99 %</b> (99)		<b>97</b> % (70)		

Kuten taulukosta 9 näkee, taitotasolla A1 ei ole yhtään LEHTI-sanatyypin sanoja. Muutoin KATTO-sanoja tämän tason sanoista on 74 % ja PANKKI-sanoja sanoista on 26 %. KATTO-typin norminmukaisuus on 90 % ja PANKKI-sanojen 86 %. Virheelliset sanat ovat KATTO-typin osalta *viikoa* ja *kukata*. PANKKI-sanojen virheellinen muoto on *säkkkin*. Näitä kaikkia sanoja on osattu taivuttaa sanatyypin vaatimalla tavalla, jos ei oteta huomioon astevaihtelua. Silti prosentuaalisesti vaikein sanatyypin tällä tasolla on PANKKI.

Taitotasolla A2 ei myöskään ole LEHTI-typin sanoja ollenkaan. KATTO-typin sanoja on tällä tasolla 46 % ja PANKKI-typin sanoja on 54 %. Tällainen jakauma on poikkeuksellinen, koska millään muulla taitotasolla edes kvalitatiivisen vaihtelun osalta jakauma ei ole sellainen, että PANKKI-typin sanoja olisi enemmän kuin KATTO-typin sanoja. Norminmukaisuus on KATTO-sanojen osalta 76 % ja PANKKI-sanojen osalta 87 %. KATTO-typin osalta virheelliset muodot ovat *viikoa*, *viikoa*, *Pekkan*, *Pekkan*, *Pekkan*, *Jukkan*, *Jukkan*, *kirkoon*, *viikona*, *kirkossa* ja *kirkossa*. Muutoin näitä kaikkia sanoja on osattu muodostaa ihan oikein, mutta astevaihtelun kannalta muodot ovat norminvastaisia. PANKKI-sanojen virheelliset muodot ovat *musiikita*, *musiikkien*, *musiikkian*, *musiikkista*, *diskoteikiin*, *karkeä* ja *pankkien*. Näistä muodoista mielenkiintoisin sanatyypin valossa on *karkeä*, koska muodolla on haettu sanaa *karkkia*. Näin ollen tässä sanassa on tehty LEHTI-sanatyypin kuuluva *i-e*-vartalovaihtelu, vaikka tämä ei kuulu tähän siihen sanatyypin, jota sana edustaa. Pääte on myös vokaaliharmonian vastainen. Muutoin nämä sanat on taivutettu oikein pois lukien astevaihtelun osuus. Tällä tasolla vaikein sanatyypin astevaihtelun valossa on siis KATTO, eikä se ole hallinnassa 80 %:n rajaan suhteutettuna.

Seuraavalla tasolla B1 jakauma on hyvin erilainen kuin aiemmalla tasolla. Tällä tasolla KATTO-tyyppin sanoja on 80 % ja PANKKI-tyyppin sanoja 20 %, eikä tälläkään tasolla ole siis LEHTI-tyyppin sanoja laisinkaan. Norminmukaisuus on KATTO-sanojen osalta 89 % ja PANKKI-sanojen osalta 95 %. KATTO-tyyppin sanojen osalta virheelliset muodot ovat *kukata, serkuuni, kukatta, tupakkalle, Kreikkassa, kreikkassa, keskiviikona, kukkoksi ja paikaan*. Näitä kaikkia on taivutettu sanatyypille tyypillisellä tavalla, mutta astevaihtelua on käytetty virheellisesti. PANKKI-tyyppin virheellinen muoto on *muusiikia*, joka on muutoin oikein, mutta siinä on yksi *u* liikaa ja astevaihtelua on käytetty virheellisesti. Tällä tasolla vaikein sanatyyppejä on siis KATTO, mutta sekin on prosentuaalisesti hallinnan puolella. B2-tasolla on ensimmäisen kerran mukana myös LEHTI-tyyppin sanoja. Muutoin KATTO-sanoja on kaikista tämän vaihtelun ja tämän tason sanoista 65 %, PANKKI-sanoja on 30 % ja LEHTI-sanoja on 5 %. Kaikki PANKKI- ja LEHTI-sanat ovat oikein, mutta KATTO-sanojen norminmukaisuusprosentti on 97. Nämä virheelliset muodot ovat *särkkän* ja *luokaa*, joita kumpaakin on taivutettu oikein muutoin kuin astevaihtelun osalta. Tällä tasolla vaikein sanatyyppejä on siis KATTO.

Taitotasolla C1 on taas kaikkia tarkastelussa mukana olevien sanatyyppejä sanoja. 81 % sanoista on KATTO-tyyppin sanoja, 15 % PANKKI-tyyppin sanoja, ja 4 % on LEHTI-tyyppin sanoja. Muutoin kaikki muodot ovat oikein, mutta PANKKI-tyyppin norminmukaisuus on 93 %. Tämä virheellinen muoto on *musiikkin*, joka on muutoin taivutettu oikein, mutta astevaihtelupari on vääränvahvuinen. Näin ollen tämän tason haastavin sanatyyppejä on PANKKI. Viimeiselläkin tasolla eli tasolla C2 on kaikkia tarkasteltavia sanatyyppejä. KATTO-sanoja sanoista on 82 %, PANKKI-sanoja on 8 %, ja LEHTI-sanoja on 10 %. Tällä tasolla poikkeuksellista on se, että PANKKI-sanoja on suhteessa niin vähän, ja niitä on jopa vähemmän kuin kyseisen tason LEHTI-sanoja. Muutoin kaikki sanat ovat oikein, mutta KATTO-sanojen norminmukaisuusprosentti on 79. Nämä virheelliset muodot ovat *kirkkon* ja *nurkaan*, joita on astevaihtelua lukuun ottamatta taivutettu sanatyypin mukaisella tavalla. Tällä tasolla tämä on siis haastavin sanatyyppejä. Jos ajatellaan 80 %:n rajaa, tämä sanatyyppejä ei ole hallinnassa enää korkeimmalla tasolla.



Kuvio 11. Kvantitatiivisen vaihtelun sanojen sanatyyppeiden norminmukaisuuden kehitys taitotasoin. Kuten kuviosta 11 näkyy, sanatyyppeiden norminmukaisuudet vaihtelevat jonkin verran. KATTO-sanojen norminmukaisuus lähtee 90 %:n tuntumasta, jonka jälkeen se laskee suhteellisen radikaalisti. Tämän jälkeen se nousee tasaisesti, kunnes se on 100 % tasolla C1, jonka jälkeen se laskee vielä radikaalisti alle 80 %:iin. PANKKI-sanojen norminmukaisuus lähtee alle 90 %:sta, jonka jälkeen se pysyy ensin melkein samana ja sitten nousee tasaisesti, kunnes se on 100 % tasolla B2. Tämän jälkeen se laskee melko paljon, jonka jälkeen se on taas 100 % viimeisellä taitotasolla. LEHTI-sanoja on vain kolmella viimeisellä tasolla, ja niiden norminmukaisuus on koko ajan 100 %.

#### 4.2.6 Sanojen syntaktisen funktion jakauma ja tarkkuus

Kuten mainitsin jo luvussa 5.1.6, syntaktisen funktion tarkastelu eroa aiemmista tätä vaihtelua käsittelevistä luvuista, koska tässä luvussa ei tarkastella sanojen rakennetta. Taulukosta 10 näkee, että kaikilla taitotasolla ei ole kaikkien lauseenjäsenten edustajia. Tasolla A1 ei ole ollenkaan predikatiiveja, eikä tasolla C1 ole ollenkaan subjekteja. Muutoin eniten on adverbiaaleja (64 %), sitten objekteja (25 %), sitten e-subjekteja (5 %), ja lopuksi sekä subjekteja että predikatiiveja on 3 %. Norminmukaisuudet vaihtelevat melkoisen paljon. Eniten oikein on objekteja (95 %), sitten adverbiaaleja (94 %), sitten predikatiiveja (93 %), sitten e-subjekteja (84 %), ja vähiten oikein on subjekteja (67 %). Prosentuaalisesti tämän

vaihtelun alaisista sanoista vaikeimpia ovat siis subjektit, mutta määrällisesti vaikeimpia ovat adverbiaalit.

Taulukko 10. Kvantitatiivisen vaihtelun sanojen syntaktisen funktion jakauma ja tarkkuus.

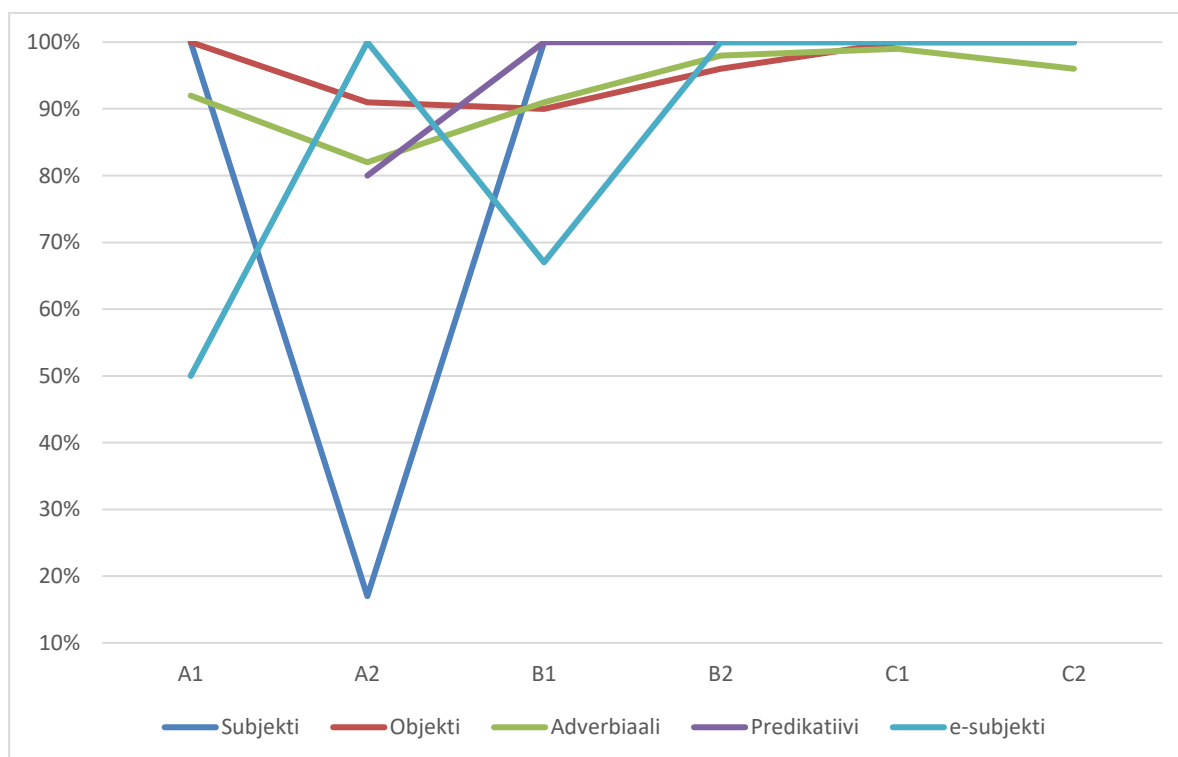
	<b>A1</b>		<b>A2</b>		<b>B1</b>		<b>B2</b>		<b>C1</b>		<b>C2</b>		Yhteensä syntaktista funktiota	
	n=27		n=100		n=100		n=100		n=100		n=72			
	määrä	NM	määrä	NM	määrä	NM	määrä	NM	määrä	NM	määrä	NM	määrä	NM
<b>Subjekti</b>	4 % (1)	<b>100 %</b> (1)	6 % (6)	<b>17 %</b> (1)	1 % (1)	<b>100 %</b> (1)	5 % (5)	<b>100 %</b> (5)	-	-	3 % (2)	<b>100 %</b> (2)	3 % (15)	<b>67 %</b> (10)
<b>Objekti</b>	37 % (10)	<b>100 %</b> (10)	34 % (34)	<b>91 %</b> (31)	20 % (20)	<b>90 %</b> (18)	24 % (24)	<b>96 %</b> (23)	20 % (20)	<b>100 %</b> (20)	26 % (19)	<b>100 %</b> (19)	25 % (127)	<b>95 %</b> (121)
<b>Adverbiaali</b>	44 % (12)	<b>92 %</b> (11)	51 % (51)	<b>82 %</b> (42)	70 % (70)	<b>91 %</b> (64)	63 % (63)	<b>98 %</b> (62)	74 % (74)	<b>99 %</b> (73)	67 % (48)	<b>96 %</b> (46)	64 % (318)	<b>94 %</b> (298)
<b>Predikatiivi</b>	-	-	5 % (5)	<b>80 %</b> (4)	3 % (3)	<b>100 %</b> (3)	4 % (4)	<b>100 %</b> (4)	1 % (1)	<b>100 %</b> (1)	1 % (1)	<b>100 %</b> (1)	3 % (14)	<b>93 %</b> (13)
<b>e-subjekti</b>	15 % (4)	<b>50 %</b> (2)	4 % (4)	<b>100 %</b> (4)	6 % (6)	<b>67 %</b> (4)	4 % (4)	<b>100 %</b> (4)	5 % (5)	<b>100 %</b> (5)	3 % (2)	<b>100 %</b> (2)	5 % (25)	<b>84 %</b> (21)
<b>Yhteensä</b>		<b>89 %</b> (24)		<b>82 %</b> (82)		<b>90 %</b> (90)		<b>98 %</b> (98)		<b>99 %</b> (99)		<b>97 %</b> (70)		

Taitotasolla A1 on eniten adverbiaaleja (44 %) ja objekteja (37 %). Paljon vähemmän on e-subjekteja (15 %) ja subjekteja (4 %). Kaikki subjektit ja objektit ovat oikein, ja adverbiaaleista oikein on 92 % ja e-subjekteista vain 50 %. Tämän tason vaikein lauseenjäsän astevaihtelun kannalta on siis e-subjekti. Taitotasolla A2 on taas eniten objekteja (34 %) ja adverbiaaleja (51 %). Tällä tasolla on ensimmäisen kerran predikatiiveja (5 %), ja myös subjekteja (6 %) ja e-subjekteja on vähemmän (4 %). Tällä tasolla kaikki e-subjektit ovat oikein, objektien norminmukaisuusprosentti on 91, adverbiaalien norminmukaisuus on 82 %, predikatiivien 80 % ja subjektien 17 %. Tämän tason vaikein lauseenjäsän on siis subjekti, joka johtuu siitä, että tämän tason virheellisistä subjekteista monet ovat miesten nimiä, joissa on genetiivissä vahva muoto. Nämä virheelliset muodot ovat jo monesti aiemmin esiintyneet *Jukkan-* ja *Pekkan-*muodot.

Taitotasolla B1 on vain 1 % subjekteja, 3 % predikatiiveja ja 6 % e-subjekteja. Eniten on siis objekteja (20 %) ja adverbiaaleja (70 %). Kaikki subjektit ja predikatiivit ovat oikein.

Adverbiaaleista oikein on 91 %, objekteista on oikein 90 %, ja vähiten oikein on e-subjekteja (67 %). Prosentuaalisesti vaikein lauseenjäsän tällä tasolla on siis e-subjekti. Tasolla B2 jakauma on hyvin pitkälti samanlainen kuin aiemmallakin tasolla. Eniten on adverbiaaleja (63 %), sitten objekteja (24 %), sitten subjekteja (5 %), ja 4 % on sekä predikatiiveja että e-subjekteja. Kaikki subjektit, predikatiivit ja e-subjektit ovat oikein. Adverbiaaleista oikein on 98 % ja objekteista 96 %. Tällä tasolla vaikein lauseenjäsän prosentuaalisesti on objekti, vaikka objektien virheellisiä muotoja onkin vain yksi.

Taitotasolla C1 on selkeästi eniten adverbiaaleja, koska niitä on jopa 74 %. Toiseksi eniten on objekteja, joita on 20 %. Vähiten on e-subjekteja (5 %) ja predikatiiveja (1 %), eikä subjekteja ole ollenkaan. Kaikissa muissa syntaktisissa funktioissa kaikki muodot ovat oikein, mutta adverbiaaleissa on yksi virheellinen muoto, jolloin norminmukaisuusprosentti on 99. Näin ollen tämä on siis tällä tasolla vaikein lauseenjäsän. Viimeisellä taitotasolla eli tasolla C2 on taas kaikkien lauseenjäsänen edustajia. Eniten on taas adverbiaaleja (67 %), sitten objekteja (26 %) ja vähiten subjekteja (3 %), e-subjekteja (3 %) ja predikatiiveja (1 %). Muutoin kaikki tämän tason muodot ovat oikein, mutta adverbiaaleissa on kaksi virheellistä muotoa, jolloin norminmukaisuusprosentti on 96. Tällä tasolla haastavin lauseenjäsän on siis adverbiaali.



Kuvio 12. Kvantitatiivisen vaihtelun sanojen norminmukaisuuden kehitys taitotasoinen tietyssä syntaktisessa funktiossa.

Kaikkien syntaktisten funktioiden sanojen norminmukaisuudet vaihtelevat melko paljon ensimmäisillä taitotasolla, kuten kuviosta 12 näkee. Subjektien norminmukaisuus on muutoin 100 %, mutta se tekee hyvin voimakkaan notkahduksen taitotasolla A2, jolloin se laskee jopa alle 20 %:n. Objektien norminmukaisuus lähtee 100 %:sta, jonka jälkeen se laskee hieman kahdella seuraavalla tasolla, josta se taas nousee 100 %:iin. Adverbiaalien norminmukaisuus lähtee suhteellisen korkealta, yli 90 %:sta. Tämän jälkeen se laskee hieman, josta se nousee tasaisesti, mutta se ei ole missään vaiheessa 100 %. Predikatiivien norminmukaisuus on ensiesiintymisen tasolla 80 %, jonka jälkeen se nousee heti 100 %:iin ja pysyy siellä. E-subjektien norminmukaisuus vaihtelee melko paljon. Ensin se lähtee 50 %:sta, josta se nousee heti 100 %:iin. Tämän jälkeen se taas laskee, mutta se ei kuitenkaan laske enää lähtötasolle. Sen jälkeen se nousee 100 %:iin, eikä se laske enää sen alle seuraavilla taitotasolla. Alle 80 %:n eli hallinnan tason alle laskevat sekä subjektisanat että e-subjektisanat. Ne ovat tämän tason alla kuitenkin vain kolmella eri taitotasolla. Kokonaisuudessaan subjektit eivät ole hallinnassa, koska niiden kokonaisnorminmukaisuusprosentti on 67.

## 5 Lopuksi

Edellisessä luvussa analysoin aineiston erikseen sekä kvalitatiivisen että kvantitatiivisen vaihtelun osalta ja jokaisella taitotasolla erikseen. Tässä luvussa vertailen molempia vaihtelutyyppejä toisiinsa ja teen yhteenvetoa.

### 5.1 Yhteenveto

$k$ :n sisältäviä astevaihtelusanoja esiintyy jokaisella taitotasolla, eli niiden ensiesiintyminen on kummankin vaihtelun osalta taitotasolla A1. Kvalitatiivisen vaihtelun alaisista muodoista oikein on kokonaisuudessaan 95 % ja kvantitatiivisen vaihtelun alaisista sanoista oikein on 93 %. Näin ollen voidaan todeta, että ainakin tällä otannalla kvantitatiivinen vaihtelu on ollut oppijoille haastavampi kuin kvalitatiivinen vaihtelu. Tällainen tulos on siinä mielessä vähän yllättävä, että kvalitatiivisessa vaihtelussa tulee osata enemmän vaihtelupareja kuin kvantitatiivisissa vaihtelussa. Vain yhden mahdollisen vaihteluparin ( $kk$  vaihtelee aina  $k$ :n kanssa) voisi ajatella olevan helpompi osata kuin monen eri vaihteluparin. Kummankin vaihtelun osalta  $k$ :n suora astevaihtelu on jo todella lähellä täyttä norminmukaisuutta ensimmäisen kerran taitotasolla B2.

Astevaihtelusanojen frekvenssi on hieman erilainen riippuen vaihtelun tyypistä. Kvalitatiivisen vaihtelun alaisten sanojen frekvenssi on hieman korkeampi kuin kvantitatiivisen vaihtelun alaisten sanojen. Kvalitatiivisen vaihtelun osalta frekvenssi liikkuu 37 ja 47 välillä, kun taas kvantitatiivisen vaihtelun sanoissa frekvenssi liikkuu 24 ja 34 välillä. Frekvenssi on laskettu tuhatta sanaa kohden. Näin ollen oppijat käyttävät enemmän kvalitatiivisen vaihtelun alaisia sanoja kuin kvantitatiivisen vaihtelun sanoja. Kvalitatiivisen vaihtelun frekvenssi on siis pienimmilläänkin isompi kuin kvantitatiivisen vaihtelun korkein frekvenssi. Frekvenssit eroavat toisistaan myös kehitykseltään. Kvalitatiivisen vaihtelun osalta korkein astevaihtelusanojen frekvenssi on taitotasolla B2, kun taas kvantitatiivisen vaihtelun osalta korkein frekvenssi on taitotasolla B1. Molemmissa vaihteluissa frekvenssit kasvavat tasaisesti korkeimmalle tasolle, jonka jälkeen frekvenssit taas laskevat ja nousevat aivan aavistuksen vielä viimeisellä tasolla. Siinä mielessä frekvenssit myös käyttäytyvät samoin, että molemmissa vaihtelutyypeissä korkeimmat frekvenssit ovat keskitasolla, ja alimmilla ja ylimmillä tasoilla ne ovat keskenään melkein samanlaiset. Frekvenssit noudattelevat muutoin korpuksen sanamääriä tietyillä taitotasoilla, mutta frekvenssin nouseminen tasolta C1 tasolle C2 siirryttäessä ei korreloi taitotasojen sanamäärien kanssa.

Frekvenssin katsotaan kertovan sujuvuudesta, mutta kuten aiemmin mainitsin, se ei ole niin yksinkertaista. Tässä tutkielmassa se kertoo enemmän siitä, kuinka paljon oppijoiden käyttämät sanat sisältävät *k:n* astevaihtelua.

Vaihteluparia *k : ø* (*joki : joen*), *k : ø (j)* (*aika : ajan*), *nk : ng* (*kenkä : kengän*) ja *k : kk* (*kirkko : kirkon*) esiintyy ensimmäisen kerran taitotasolla A1, eli näitä esiintyy jo kielen oppimisen alkuvaiheessa. Taitotasolla B1 esiintyy ensimmäisen kerran vaihteluparia *k : j* (*kärki : kärjen*) ja tasolla B2 esiintyy ensimmäisen kerran vaihteluparia *k : v* (*luku : luvun*). Jälkimmäisellä perustason taitotasolla esiintyy siis jo kaikkia vaihtelupareja. Vertailtaessa kvalitatiivista ja kvantitatiivista vaihtelua huomataan, että kvantitatiivinen vaihtelu on vaikein vaihtelupari taitotasolla A2, B2 ja C2.

Kokonaistarkkuusprosentiltaan vaikein vaihtelupari otannassa on kuitenkin *k : j* -pari, mutta näitä on vain 1 % koko otannan sanoista, joten vaikeimmaksi vaihtelupariksi voidaan myös mieltää *nk : ng* -pari, koska näiden kokonaistarkkuusprosentti on 89 % ja koko otannasta näitä sanoja on 31 %. *k : j* -vaihtelupari on hallinnassa taitotasolla B2, mutta kokonaisuudessaan sen tarkkuusprosentti ei ylitä 80 %:n rajaa. *nk : ng* -pari on hallinnassa ensimmäisen kerran taitotasolla A2, ja sen kokonaistarkkuusprosentti osoittaa, että se on hallinnassa. Helpoin vaihtelupari on *k : v*, koska kaikki tätä vaihteluparia sisältävät sanat ovat tässä otannassa oikein. Tätä vaihteluparia on kylläkin vain kolmella viimeisellä tasolla ja vain viidessä eri sanassa suomen kielessä.

Jakaumana tarkastellaan vaihteluparien määrää suhteessa koko otannan sanoihin, jolloin kvantitatiivisen vaihtelun sanojen jakaumasta ei oikein voi sanoa mitään. Kvalitatiivisen vaihtelun sanojen jakauma vaihteluparien suhteen on suhteellisen tasainen *k : ø* -, *k : ø (j)* ja *nk : ng* -parien välillä (30–36 %) ja kaikkein vähiten on *k : j* - ja *k : v* -pareja (1–2 %). Jos tätä peilaa 1 000 yleisimmän sanan astevaihteluesiintymiin (Karlsson 1983, 328), *nk : ng* -vaihtelupareja on suhteessa enemmän tässä otannassa, vaikka molemmat kadolliset muodot laskettaisiinkin yhdeksi vaihtelupariksi. Tästä ei kuitenkaan voi vetää ihan näin suoria johtopäätöksiä, koska laskutavat ovat hyvin erilaiset.

Otannan sanoista genetiivissä, partitiivissa, inessiivissä, elatiivissa, illatiivissa, adessiivissa ja ablatiivissa olevia sanoja esiintyy ensimmäisen kerran taitotasolla A1. Translatiivissa, allatiivissa ja essiivissä esiintyy sanoja ensimmäisen kerran tasolla A2.

Possessiivisuffiksillisia muotoja esiintyy ensimmäisen kerran taitotasolla A1, ja liitepartikkelin sisältäviä muotoja on ensimmäisen kerran taitotasolla A2. Kaikkia sijamuotoja

(pois lukien nominatiivi ja vajaakäyttöiset sijat) esiintyy siis jo ensimmäisillä taitotasoilla. Vaikein sijamuoto astevaihtelun kannalta on molempien vaihteluiden osalta translatiivi, vaikkakin näitä on molemmissa vain noin 1 % kaikista sanoista. Eniten kummankin vaihtelun alaisista sanoista sanoja on genetiivissä, partitiivissa ja inessiivissä (16–27 %). Näistä sijamuodoista vaikein kummankin vaihtelun osalta on genetiivi. Kvalitatiivisen vaihtelun osalta helpoimmat sijamuodot ovat ablatiivi, allatiivi ja essiivi ja kvantitatiivisen vaihtelun osalta adessiivi ja ablatiivi, koska näissä on kaikki muodot oikein. Kvalitatiivisen vaihtelun osalta hallinnassa ei kokonaisuudessaan ole adessiivi eikä translatiivi. Näitä kaikkia tosin on vain muutamia prosentteja otannassa, joka voi selittää korkeaa tarkkuusprosenttia. Kvantitatiivisen vaihtelun osalta kaikki sijamuodot ovat hallinnassa.

Virhetyyppejä tarkasteltaessa voi huomata, että kvalitatiivisen vaihtelun osalta on paljon yleisempää, että astevaihtelua ei käytetä sellaisissa muodoissa, jossa sitä tulisi käyttää. Näin päin virheellisiä muotoja on 22 ja toisin päin vain viisi. Kvantitatiivisen vaihtelun osalta virhetyyppien jakauma on paljon tasaisempi. Astevaihtelun käyttämättömyyttä on 20 sanassa ja tarpeetonta käyttöä on 16 sanassa.

Kvantitatiivisen vaihtelun alaiset sanat ovat pidempiä kuin kvalitatiivisen vaihtelun alaiset sanat, koska jälkimmäisen vaihtelun sanoissa ei ole yhtäkään sellaista muotoa, jossa olisi kuusi tavua. 2–4-tavuisia sanoja on ensimmäisen kerran otannassa jo ensimmäisellä taitotasolla, viisitavuisia muotoja on tasolla A2 ja kuusitavuisia muotoja on tasolla B1. Eniten kvalitatiivisen vaihtelun alaisista sanoista on kaksitavuisia muotoja ja kvantitatiivisen vaihtelun sanoista on eniten kolmitavuisia sanoja. Yli neljätavuisia sanoja otannassa on kaikkein vähiten molempien vaihteluiden osalta. Kvalitatiivisen vaihtelun alaisissa sanoissa eniten virheellisiä muotoja on kolmetavuisissa sanoissa, ja kvantitatiivisen vaihtelun sanoissa eniten virheellisiä muotoja on kaksitavuisissa sanoissa, eli lyhimät sanat ovat tämän vaihtelun osalta vaikeimpia. Kaiken pituiset sanat ovat otannassa kokonaisuustarkkuusprosenttiltaan hallinnassa, vaikka yksittäisillä taitotasoilla voitiinkin jäädä alle 80 %:n jonkin tavumäärän osalta.

Kvalitatiivisen vaihtelun sanoissa kaikkein eniten virheellisiä muotoja on sellaisissa sanoissa, joihin lisätään päätte, joka on muotoa CCV, jolloin edellinen tavu muuttuu joko muotoon CVC tai CV. Kvantitatiivisen vaihtelun sanoissa puolestaan eniten virheellisiä muotoja on sellaisissa sanoissa, joihin lisätään päätte, joka on muotoa C, jolloin viimeinen tavu muuttuu muotoon CVC.

Kvalitatiivisen vaihtelun osalta kaikki tarkastellut sanatyypit tekevät ensiesiintymisensä jo ensimmäisellä taitotasolla. Kvantitatiivisen vaihtelun osalta LEHTI-sanatyyppi esiintyy ensimmäisen kerran vasta taitotasolla B2. Molempien vaihteluiden osalta kaikkein eniten on sanoja, joiden sanatyyppi on KATTO (70–72 %), sitten PANKKI (22–26 %) ja vähiten LEHTI (3–6 %). Prosentuaalisesti kvalitatiivisen vaihtelun sanojen sanatyypeistä vaikein tyyppi on LEHTI ja kvantitatiivisen vaihtelun sanojen sanatyypeistä PANKKI. Tosin LEHTI-tyypin sanoja on niin vähän, että myös kvalitatiivisen vaihtelun osalta vaikeimmaksi sanatyypiksi voisi mieltää PANKKI-tyypin. LEHTI-sanatyyppiin kuuluu vokaalivaihtelu *i-e*, joka on otannan virheellisissä muodoissa sisällytetty kahteen muotoon ja yhdestä se puuttuu. Mukana on myös yksi muoto, jossa astevaihtelu on oikein, mutta vartalovaihtelu puuttuu. Kokonaisuustarkkuusprosentiltaan molempien vaihteluiden osalta kaikki sanatyypit ovat hallinnassa, tosin joillakin taitotasoilla jotkin sanatyypit jäävät alle 80 %:n.

Jo ensimmäisellä taitotasolla on astevaihtelusanoja kaikissa syntaktisissa funktioissa. Kaikkein eniten otannan sanoista on adverbiaaleja ja toiseksi eniten objekteja. Kvalitatiivisen vaihtelun osalta adverbiaaleissa on eniten virheellisiä muotoja ja kvantitatiivisen vaihtelun osalta eniten virheellisiä muotoja on ehkä vähän yllättävästikin subjekteissa. Kumpikaan näistä ei ole hallinnassa kokonaistarkkuusprosentin mukaan. Joillakin yksittäisillä taitotasoilla jotkin muutkaan lauseenjäsenet eivät vielä ole hallinnassa.

## 5.2 Päätelmät

Tutkimuskysymykseni siis olivat:

1. Miten astevaihtelun käytön tarkkuus korreloi oppijan kielitaitotason kanssa?
2. Miten astevaihtelun alaisten sanojen osuus muuttuu taitotasojen välillä, ja mikä niiden osuus on kaikista korpuksen substantiiveista?
3. Miten käytettyjen astevaihtelun alaisten sanojen rakenne ja syntaktinen funktio vaikuttavat eri muotoihin eri taitotasoilla?
4. Millaisia eroja on eri vaihteluparien oppimisen välillä?

Suoran *k*:n astevaihtelun tarkkuus korreloi oppijan kielitaitotason kanssa niin, että ensimmäisellä tasolla tarkkuus on 82,5–89 %, toisella tasolla se on 82–87 %, kolmannella 90–94 %, neljännellä 98–100 %, viidennellä 99 % ja kuudennella 97–100 %. Tarkkuus ei siis

nouse lineaarisesti tasolta toiselle siirryttäessä, mutta se kyllä nousee keskimäärin sitä paremmaksi, mitä korkeammasta taitotasosta on kyse.

$k:n$  suoran astevaihtelun alaisten sanojen osuus muuttuu taitotasolta toiselle siirryttäessä niin, että ensin määrä kasvaa tasaisesti saavuttaen huippunsa tasoilla B1 ja B2, jonka jälkeen se laskee melkein lähtötasolle, jonka jälkeen se nousee vielä vähän viimeisellä taitotasolla. Viimeisellä taitotasolla ei kuitenkaan olla enää lähellä huipputasoa. Osuus kaikista substantiiveista vaihtelee 24 sanasta tuhatta sanaa kohden noin 47 sanaan tuhatta sanaa kohden.

Käytettyjen sanojen rakenne ja syntaktinen funktio vaikuttavat käytettyihin muotoihin monin eri tavoin eri taitotasolla. Tämän tutkimuksen mukaan suomenoppijalle vaikein sanamuoto on sellainen sana, jossa on  $nk : ng$  -vaihtelupari, sana on translatiivissa tai genetiivissä, sana on kaksi tai kolmitavuinen ja siihen lisätään joko CCV-pääte tai C-pääte, se on sanatyypiltään PANKKI, ja lauseenjäsenenä se on adverbiaali, ja todennäköisesti siinä on käytetty astevaihtelua, vaikka ei olisi kuulunut. Keskimäärin tällaiset asiat sisältävä muoto on siis ollut haastavin, koska nämä ovat haastavimmat tarkasteltavat muuttujat. Tällainen sana voisi olla siis esimerkiksi kaupungin tai kaupungiksi, ja virheellinen muoto olisi *kaupunkin* tai *kaupunkiksi*.

Eri vaihteluparien oppimisen välillä on myös paljon eroja eri taitotasolla. Vaikein vaihtelupari on näkökulmasta riippuen joko  $k : j$  - tai  $nk : ng$  -pari. Joillakin taitotasolla vaikein vaihtelupari on myös  $kk : k$ , joka on siis kvantitatiivisen vaihtelun ainoa vaihtelupari. Kokonaisuudessaan kvantitatiivinen vaihtelu on hieman haastavampi kuin kvalitatiivinen vaihtelu.

Kaiken kaikkiaan voin todeta, että  $k:n$  suora astevaihtelu on suhteellisen hyvin hallinnassa suomen oppijoilla jo kielen oppimisen alkuvaiheesta alkaen, koska millään tasolla kokonaistarkkuusprosentti ei ole alle 80 %. Joidenkin muuttujien osalta joillakin taitotasolla tarkkuusprosentti kyllä on alle 80. Kaikkien tarkasteltavien muuttujien valossa löysin eroja sanamuodoissa eri taitotasolla, ja vain harvan muuttujan suhteen kaikki muodot olivat oikein jokaisella taitotasolla. Näin ollen voi siis sanoa, että sanan rakenne vaikuttaa astevaihtelun norminmukaisuuteen. Tutkielman tulokset kertovat, että astevaihtelun oppimiseen vaikuttavat monet tekijät, mutta pääosin se kuitenkin alkaa olla hallinnassa jo lähes kokonaan taitotasolla B2.

Koska tässä tutkielmassa tutkin sanoja ja niiden rakennetta, ei voida poissulkea puhtaiden kirjoitusvirheiden vaikutusta tuloksiin. Olen siis olettanut, että jokainen virheellinen muoto on johtunut astevaihteluvirheestä, vaikka todellisuudessa on hyvin mahdollista, että kyseessä on ollut vain kirjoitusvirhe, ja oikeasti oppija olisi voinut osata muodostaa oikean muodon. Osaan virheistä on voinut vaikuttaa myös tietämättömyys erilaisten sanojen taivutuksesta. Otannassa on mukana esimerkiksi monta nimeä, joissa on astevaihteluvirhe, jolloin virheellisen muodon on voinut aiheuttaa se, ettei oppija ole tiennyt nimienkin yleensä olevat astevaihtelun alaisia. Mukana on myös tekstejä samoilta oppijoilta, ja sanat voivat olla myös samoista teksteistä, jolloin virheellinen muoto on voinut sen takia toistua.

Tätä aihetta voisi tutkia lisää monellakin eri tavalla. Esimerkiksi myös muita astevaihtelussa olevia klusiileja sisältäviä sanoja voisi tutkia samoin ja verrata näitä vaihteluja keskenään. Tässä tutkielmassa ei myöskään tarkasteltu käänteistä astevaihtelua, jota voisi myös olla mielenkiintoista tutkia ja verrata suoran astevaihtelun sanoihin. Myös sellainen tutkimusasetelma, jossa tutkitaan saman oppijan astevaihtelusanon kehitystä voisi olla hedelmällinen, jolloin tutkimuksen fokuksessa olisi enemminkin astevaihtelun kehitys tietyllä oppijalla ja siihen vaikuttavat tekijät. Tähän voisi tietysti yhdistää myös opetustavan vaikutuksen, koska eri oppikirjoissa astevaihtelua opetetaan hieman eri tavoin. Olisi esimerkiksi hyvin mielenkiintoista tietää, olisiko astevaihtelua parempi opettaa umpi- ja avotavuihin keskittyen vai opettamalla säännöt sanamuotoihin, joissa astevaihtelua esiintyy.

## Lähteet

### Aineslähteet

ICLFI = Jantunen, J., Brunni, S., & Oulun yliopisto, suomen kielen oppiaine (2013).  
Kansainvälinen oppijansuomen korpus [tekstikorpus]. Kielipankki. Saatavissa:  
<http://urn.fi/urn:nbn:fi:lb-20140730163>

### Muut lähteet

- Aalto, Eija – Latomaa, Sirkku – Suni, Minna 1996. *Kolmen vuosikymmenen satoa – Suomi toisena ja vieraana kielenä -alan bibliografia*. Soveltavan kielentutkimuksen keskus. Jyväskylän yliopisto.
- Aalto, Eija 1992. Alkeisoppijan strategioita suomen kielen taivutuksen prosessoinnissa. *Acquisition of Language – Acquisition of Culture*. s.7–15. AFinLa Yearbook 1992 (Suomen soveltavan kielitieteen yhdistyksen julkaisuja 50).
- Alanen, Riikka – Martin, Maisa 2011. Oppijankieli – kehitystä ja virheitä. s.33–48. *Kieltä tutkimassa: tutkielman laatijan opas*. Toim. Paula Kalaja, Riikka Alanen ja Hannele Dufva. Finn Lectura.
- Bulté, Bram – Housen, Alex 2012. Defining and operationalizing L2 complexity. *Dimensions of L2 Performance and Proficiency. Complexity, Accuracy and Fluency in SLA*. s. 21–46. Toim. Alex Housen, Folkert Kuiken ja Ineke Vedder. *Language Learning & Language Teaching* 32. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Council of Europe 2023. *Democracy and Human Dignity: Linguistic Integration of Adult Migrants (LIAM): Themes: EVK ja profiilit*. [Viitattu 4.2.2023.] Saatavissa: <https://www.coe.int/fi/web/lang-migrants/cefr-and-profiles>
- Dahl, Östen 2008. Kuinka eksoottinen kieli suomi on?. s.545–559. *Virittäjä* 4/2008.
- Eränummi, Saana 2021. *Kieltokonstruktion frekvenssin, tarkkuuden ja distribuution kehitys suomen kielen alkeisoppijoiden teksteissä*. Turun yliopisto. Humanistinen tiedekunta. Kieli- ja käänntieteen laitos.
- Gehring, Sonja – Heinzmann, Sonja 2020. *Suomen mestari 1. Suomen kielen oppikirja aikuisille*. Uudistetun painoksen 1. painos. Finn Lectura.
- Hazāne, Juliana 2009. *Latvialaiset opiskelijat suomen nimitystaitavutuksen oppijoina*. Pro gradu –tutkielma. Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.

- Heikkinen, Vesa – Lehtinen, Outi – Lounela, Mikko 2001. Kuvia kirjoitetusta suomesta. *Kielikello* 3/2001. [Viitattu 4.2.2023.] Saatavissa: <https://www.kielikello.fi/-/kuvia-kirjoitetusta-suomesta>
- Housen, Alex – Kuiken, Folkert – Vedder Ineke 2012. Complexity, accuracy and fluency: Definitions, measurement and research. *Dimensions of L2 Performance and Proficiency. Complexity, Accuracy and Fluency in SLA*. s. 1–20. Toim. Alex Housen, Folkert Kuiken ja Ineke Vedder. *Language Learning & Language Teaching* 32. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Huida, Marikki 2019. *Nesessiivirakenteet ja adjektiivit suomenoppijoiden B1-tason mielipidekirjoituksissa ja -keskusteluissa*. Pro gradu -tutkielma. Turun yliopisto. Humanistinen tiedekunta. Kieli- ja käännöstieteiden laitos.
- Jaakkola, Hanna 1997. *Kielitieto kielitaitoon pyrittäessä. Vieraiden kielten opettajien käsityksiä kielioopin oppimisesta ja opettamisesta*. Jyväskylä studies in education, psychology and social research 128. Jyväskylän yliopisto.
- Jantunen, Jarmo Harri – Pirkola Silja 2015. Oppijansuomen sähköiset tutkimusaineistot – Nykytilanne. s.88–103. *Virittäjä* 1/2015.
- Jantunen, Jarmo Harri 2011. Kansainvälinen oppijansuomen korpus (ICLFI): typologia, taustamuuttajat ja annotointi. s.86–105. *Lähivördlusi/Lähivertailuja* 21. DOI:10.5128/LV21.04
- Jussilainen, Salla 2019. *Objektin rakennetyypit ja sijanvalinta keskitason S2-oppijoiden YKI-testivastauksissa*. Turun yliopisto. Humanistinen tiedekunta. Kieli- ja käännöstieteiden laitos.
- Jyväskylän yliopisto 2012. *Cefling – Linguistic Basis of the Common European Framework for L2 English and L2 Finnish*. [Viitattu 6.1.2023] Saatavissa: <https://www.jyu.fi/hytk/fi/laitokset/kivi/tutkimus/hankkeet/paattyneet-tutkimushankkeet/cefling/suom> .
- Jyväskylän yliopisto. *Topling – toisen kielen oppimisen polut*. [Viitattu 6.1.2023] Saatavissa: <https://www.jyu.fi/hytk/fi/laitokset/kivi/tutkimus/hankkeet/paattyneet-tutkimushankkeet/topling/suom> .
- Kajander, Mikko 2013. *Suomen eksistentiaalilause toisen kielen oppimisen polulla*. Jyväskylä Studies in Humanities 220. Jyväskylän yliopisto 2013.
- Karlsson, Fred 1981. *Suomi vieraana kielenä*. Werner Söderström osakeyhtiö.
- 1982. Kieliteorian relevanssi suomen kielen opetukselle. *Suomi vieraana kielenä*. s.89–114. Toim. Fred Karlsson. Werner Söderström Osakeyhtiö.

- 1983. *Suomen kielen äänne- ja muotorakenne*. Werner Söderström Osakeyhtiö.
- 2009. *Suomen peruskielioppi*. Suomalaisen kirjallisuuden seura. 4., laajennettu painos.
- Kielitaidon tasojen kuvausasteikko 2012. *Aikuisten maahanmuuttajien kotoutumiskoulutuksen opetussuunnitelman perusteet 2012*. [Viitattu: 5.2.2023] Saatavissa: <https://eperusteet.opintopolku.fi/#/fi/kotoutumiskoulutus/7675672/tekstikappale/7789547>.
- Kielitoimiston ohjepankki 2024. *Nimet: Henkilönnimet: Nimien taivutus: sukunimet*. [Viitattu 20.3.2024] Saatavissa: <https://kielitoimistonohjepankki.fi/ohje/nimien-taivutus-sukunimet/>.
- Lehtinen, Kristiina 1998. *Astevaihtelu neljän seitsemäsluokkalaisten venäläisen maahanmuuttajan välikielessä*. Pro gradu -tutkielma. Turun yliopiston suomen kielen laitos.
- Lintunen, Pekka – Pietilä, Päivi 2014. *Kuinka kieltä opitaan: opas vieraan kielen opettajalle ja opiskelijalle*. Gaudeamus Oy.
- Maahanmuuttovirasto 2022. *Luvat ja kansalaisuus: Suomen kansalaisuus: Kansalaisuushakemus: Kielitaito*. [Viitattu 18.12.2022.] Saatavissa: <https://migri.fi/kielitaito>.
- Malin, Essi 2012. *Suomi toisena kielenä -oppijoiden sanaston kehittyminen taitotasolta toiselle siirryttäessä*. Pro gradu -tutkielma. Jyväskylän yliopiston kielten laitos.
- Martin, Maisa 1990. Suomi toisena kielenä: esimerkkejä taivutusmuotojen tuottamisesta. *Foreign Language Comprehension and Production – Vieraan kielen ymmärtäminen ja tuottaminen*. s.89–102. AFinLa Yearbook 1990 (Suomen soveltavan kielitieteen yhdistyksen julkaisuja 48).
- 1995. *The map and the rope. Finnish nominal inflection as a learning target*. *Studia Philologica Jyväskyläensia* 38. Jyväskylä: University of Jyväskylä.
- 2022. Käyttötaajuus ja tarkkuus toisen kielen kehityksessä. s.81–102. *Soveltavan kielitieteilijän sormenjälkiä etsimässä: Kielen rakenteet ja niiden käyttäjät Kirsti Siitosen tutkimuksellisina kiintopisteinä*. Toim. Lotta Aarikka, Katri Priiki ja Ilmari Ivaska. AFinLA-teema nro 14.
- Martin, Maisa – Mustonen, Sanna – Reiman, Nina – Seilonen, Marja 2010. On Becoming an Independent User. s.57–79. *Communicative proficiency and linguistic development, intersections between SLA and language testing research*. Toim. Inge Bartning, Maisa Martin ja Ineke Vedder. European Second Language Association. EUROSLA Monographs, 1.

- Mátyás, Emese 1999. Suomi toisena kielenä -alan historiaa ja nykypäivää. s.19–37. *Yhdeksän tutkielmaa suomesta toisena ja vieraana kielenä*. Toim. Maisa Martin ja Kaarlo Voionmaa. Jyväskylän yliopisto.
- Miestamo, Matti 2006. Suomi maailman kielten joukossa eli mikä suomen rakenteessa onkaan erityistä. *Kielioppi koulussa*. s.43–54. Toim. Minna Harmanen ja Mari Siironen. Äidinkielen opettajain liiton vuosikirja nro 50. Äidinkielen opettajain liitto. Helsinki.
- Pajunen, Anneli 2010. Sanojen synteettisyys suomen kielessä. *Virittäjä* 4/2010. s. 481–501
- Pöyhönen, Sari – Tarnanen, Mirja 2011. Maahanmuuttajien suomen kielen taidon riittävyys ja työllistymisen mahdollisuudet. s.139–152. *Puhe ja kieli* 4/2011. Saatavissa: <https://journal.fi/pk/article/view/4750> .
- Seilonen, Marja 2013. *Epäsuora henkilöön viittaaminen oppijansuomessa*. Jyväskylä studies in humanities 197. Jyväskylän yliopisto.
- Suni, Minna 2012. The impact of Finno-Ugric languages in second language research. s.407–438. *Lähivördlusi/Lähivertailuja* 22. DOI:10.5128/LV22.14.
- Tieteen termipankki 2022. *Kielitiede: agglutinoiva kieli*. [Viitattu 17.12.2022] Saatavissa: [https://tieteentermipankki.fi/wiki/Kielitiede:agglutinoiva\\_kieli](https://tieteentermipankki.fi/wiki/Kielitiede:agglutinoiva_kieli) .
- Tikkanen, Laura – Tähtinen, Noora 2017. *Ahaa! 1 Suomen kielen oppikirja aikuisille*. 1.–5. painos. Finn Lectura.
- Tilastokeskus 2024. *Maahanmuutto ennätyskellisen korkealla viime vuonna*. [Viitattu 5.4.2024] Saatavissa: <https://www.stat.fi/uutinen/maahanmuutto-ennatyskellisen-korkealla-viime-vuonna> .
- Vehkanen, Marjut 2015. *Kieliopista kommunikaatioon. Suomi toisena ja vieraana kielenä - oppikirjat vuosina 1866–1953*. Helsingin yliopisto. [Viitattu 10.3.2024] Saatavissa: <http://urn.fi/URN:ISBN:978-951-51-1556-0>.
- Vetikko, Melina 2021. *k:n astevaihtelun kehitys A1–C2-taitotasoilla suomi toisena kielenä - oppijoilla*. Turun yliopisto. Humanistinen tiedekunta. Kieli- ja käännöstieteiden laitos. Kandidaatintutkielma.
- VISK = Auli Hakulinen, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen ja Irja Alho 2004: *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seura. Verkkoersio. [Viitattu 17.12.2022.] Saatavissa: <http://scripta.kotus.fi/visk> URN:ISBN:978-952-5446-35-
- White, Leila 2019. *Suomen kielioppia ulkomaalaisille*. 9. painos. Finn Lectura.

Winke, Paula – Brunfaut, Tineke 2020. Perspectives on “Knowing” a Second Language – What Are We Seeking to Measure?. s. 1–8. *The Routledge Handbook of Second Language Acquisition and Language Testing*. Toim. Paula Winke ja Tineke Brunfaut. Routledge.