

**SAADA-verbin fraseologiaa: vertaileva korpustutkimus oppijan-
ja natiivisuomesta**

Maisterintutkielma

Jenny Tarvainen

Suomen kieli

Kieli- ja viestintätieteiden laitos

Jyväskylän yliopisto

2018

JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanistis-yhteiskuntatieteellinen tiedekunta	Laitos – Department Kieli- ja viestintätieteiden laitos
Tekijä – Author Jenny Tarvainen	
Työn nimi – Title SAADA-verbin fraseologiaa: vertaileva korpustutkimus oppijan- ja natiivisuomesta.	
Oppiaine – Subject Suomen kieli	Työn laji – Level Maisterintutkielma
Aika – Month and year Heinäkuu 2018	Sivumäärä – Number of pages 120 sivua + liitteet
Tiivistelmä – Abstract <p>Tutkimuksessa vertaillaan oppijansuomen ja natiivisuomen fraseologiaa piirteitä SAADA-verbin avulla. Ensin vertaillaan, missä merkityksissä SAADA-verbi esiintyy taajimmin oppijansuomen ja natiivisuomen aineistoissa, minkä jälkeen oppijansuomen aineiston yleisimmästä merkityksestä ja natiivisuomen vastaavasta merkityksestä tarkastellaan kyseisen merkityksen morfologista ja semanttista primingia. Lopuksi tilastollisia menetelmiä hyödyntäen selvitetään, onko muotoa (morfologinen priming) ja merkitystä (semanttinen priming) edustavilla elementeillä korrelaatiota.</p> <p>Työ sijoittuu kontekstuaalisen semantiikan ja fraseologian viitekehykseen, eli kieltä tutkitaan kontekstilähtöisesti. Tutkimusmenetelminä käytetään kontrastiivista oppijankielen analyysia ja korpuspohjaista tutkimusta. Lisäksi työssä hyödynnetään sekä laadullisia (verbin sisäinen ja kontekstuaalinen luokittelu) että tilastollisia analyysimenetelmiä (X^2-riippumattomuustesti ja z-testi). Oppijansuomen aineisto on kerätty Kansainvälisen oppijansuomen korpuksen (ICLFI) taitotasolta B2–C2 ja natiivisuomen aineisto Käännösuomen korpuksen (KSK) alkuperäissuomen aineistosta.</p> <p>Tutkimus osoittaa, että oppijan- ja natiivisuomessa käytetään verbiä SAADA eri tavoin: Oppijansuomessa taajin merkitys on Kielitoimiston sanakirjan (s. v. <i>saada</i>) luokituksen mukaisesti ‘tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkun tai jonkin omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jokin tulee’ ja natiivisuomen puolestaan ‘onnistua, kyetä tekemään tai aiheuttamaan jotakin tai saattamaan tekeminen päätökseen’. Myös morfologinen priming eroaa aineistojen välillä, sillä oppijansuomessa SAADA-verbin taajimman merkityksen taajimmat sananmuodot ovat <i>saa</i> ja <i>saada</i>, kun natiivisuomen vastaavan merkityksen taajimmat sananmuodot ovat <i>saanut</i> ja <i>sai</i>. Sen sijaan semanttisessa primingissa ei ollut juuri eroja, sillä molemmissa aineistoissa merkitysluokat ‘ihminen’, ‘aika, ikä’ ja ‘määrä, intensiteetti’ ylsivät neljän taajimman joukkoon. Oppijansuomessa kolmannen sijan ‘määrä, intensiteetti’ -luokan kanssa jakaa ‘raha’ ja oppijansuomessa ‘relaationaalisuus’. Molemmissa aineistoissa morfologisen ja semanttisen primingin välillä on tilastollinen yhteys, mutta oppijansuomessa yhteys on huomattavasti vahvempaa. Korrelaatio eroaa myös laadullisesti, mikä kertoo myös morfologisen primingin eroista.</p> <p>Tutkimustulokset täydentävät fraseologista leksikon kuvausta SAADA-verbin osalta, mutta myös tarkentavat käsitystämme oppijan- ja natiivisuomen eroista. Teoreettisesti tutkimustulokset tukevat ajatusta, jonka mukaan sananmuodoilla polyseemisen sanan eri merkityksillä on erilaisia kontekstuaalisia piirteitä, minkä vuoksi niitä pitäisi tutkia omina leksikaalisina yksikköinä. Lisäksi tutkimus antoi rohkaisevia menetelmällisiä tuloksia muodon ja merkityksen välisen suhteen tilastollisesta tarkastelusta.</p>	
Asiasanat – Keywords fraseologia, korpustutkimus, kontekstuaalisuus, kontrastiivinen tutkimus, vieras kieli, äidinkieli	
Säilytyspaikka – Depository	
Muita tietoja – Additional information	

SISÄLLYS

1 JOHDANTO	1
1.1 Tutkimuksen lähtökohdat	1
1.2 SAADA-verbi tutkimuskohteena	4
2 FRASEOLOGINEN VIITEKEHYS	7
2.1 Kontekstuaalinen semantiikka	7
2.1.1 Lähtökohdat ja periaatteet	8
2.1.2 Kontekstuaalinen valinta ja rajoitukset	10
2.2 Fraseologia	13
2.2.1 Fraseologia tutkimusalana	13
2.2.2 Fraseologinen yksikkö	15
2.2.3 Leksikaalinen priming	21
2.2.4 Morfologinen priming	24
2.2.5 Semanttinen priming	26
3 OPPIJANKIELI JA FRASEOLOGIA	29
3.1 Oppijankieli	29
3.1.1 Oppijankielen määritelmä ja oppimiskontekstit	29
3.1.2 Oppijankieli universaalina kielivarianttina	32
3.2 Fraseologia oppijankielessä	35
3.3 Morfologinen ja semanttinen priming oppijansuomessa	39
4 TUTKIMUSAINEISTOT	40
4.1 Korpuksat tutkimusaineistona	40
4.2 Kansainvälinen oppijansuomen korpus (ICLFI) ja Käännössuomen korpus (KSK)	41
4.3 Aineistojen vertailukelpoisuus	43
5 TUTKIMUSMETODIT	44
5.1 Kontrastiivinen oppijankielen analyysi (CIA)	44
5.2 Korpuspohjainen tutkimusmenetelmä	48
5.3 Intuition hyödyntäminen korpustutkimuksessa	51
5.4 Tilastolliset menetelmät	52

6 ANALYYSI	54
6.1 SAADA-verbin polyseemia	54
6.1.1 Merkitysluokittelun periaatteet	54
6.1.2 SAADA-verbin merkitykset oppijan- ja natiivisuomessa	67
6.2 SAADA-verbin morfologinen priming oppijan- ja natiivisuomessa	74
6.3 SAADA-verbin semanttinen priming	80
6.3.1 Semanttisen luokittelun periaatteet	80
6.3.2 SAADA-verbin semanttinen priming oppijan- ja natiivisuomessa	89
6.4 Morfologisen ja semanttisen primingin korrelaatio oppijan- ja natiivisuomessa	99
6.4.1 Luokkien yhdistämisen periaatteet ristiintaulukoinnissa	99
6.4.2 Korrelaatio morfologisen ja semanttisen primingin välillä oppijansuomessa	102
6.4.3 Korrelaatio morfologisen ja semanttisen primingin välillä natiivisuomessa	105
6.4.4 Korrelaatioiden laadullinen vertailu oppijan- ja natiivisuomessa	107
7 PÄÄTÄNTÖ	109
7.1 Tutkimustulosten yhteenveto	109
7.2 Tulokset osana traditiota	111
7.3 Tutkimuksen onnistumisen arviointi	113
7.4 Tutkimustulosten soveltamisala ja jatkotutkimus	114

LÄHTEET

LIITTEET

- Kuvio 1. Tutkimusasetelma.
- Kuvio 2. Syntagmaattinen ja paradigmaattinen ulottuvuus.
- Kuvio 3. Suomeksi mukailtu malli uudistetusta CIA-menetelmästä.
- Kuvio 4. Oppijankielen aineiston jakautuminen merkitysluokkiin (513 virkettä).
- Kuvio 5. Natiivisuomen aineiston jakautuminen merkitysluokkiin (813 virkettä).
- Kuvio 6. Oppijansuomen (513 virkettä) ja natiivisuomen (813 virkettä) merkitysluokkien vertailu järjestettynä tilastollisen merkitsevyyden mukaan.
- Kuvio 7. SAADA-verbin sananmuodot (n = 144) oppijansuomen aineiston merkitysluokasta 1.
- Kuvio 8. SAADA-verbin sananmuodot (n = 123) natiivisuomen aineistosta merkitysluokasta 1.
- Kuvio 9. Oppijansuomen (144 sananmuotoa) ja natiivisuomen (123 sananmuotoa) SAADA-verbin sananmuotojen vertailu järjestettynä tilastollisen merkitsevyyden mukaan.
- Kuvio 10. Oppijansuomen aineiston semanttiset luokat merkitysluokasta 1 (n = 475).
- Kuvio 11. Natiivisuomen aineiston semanttiset luokat merkitysluokasta 1 (n = 373).
- Kuvio 12. Oppijansuomen (475 sanetta) ja natiivisuomen (373 sanetta) semanttisten luokkien vertailu järjestettynä tilastollisen merkitsevyyden mukaan.

1 JOHDANTO

1.1 Tutkimuksen lähtökohdat

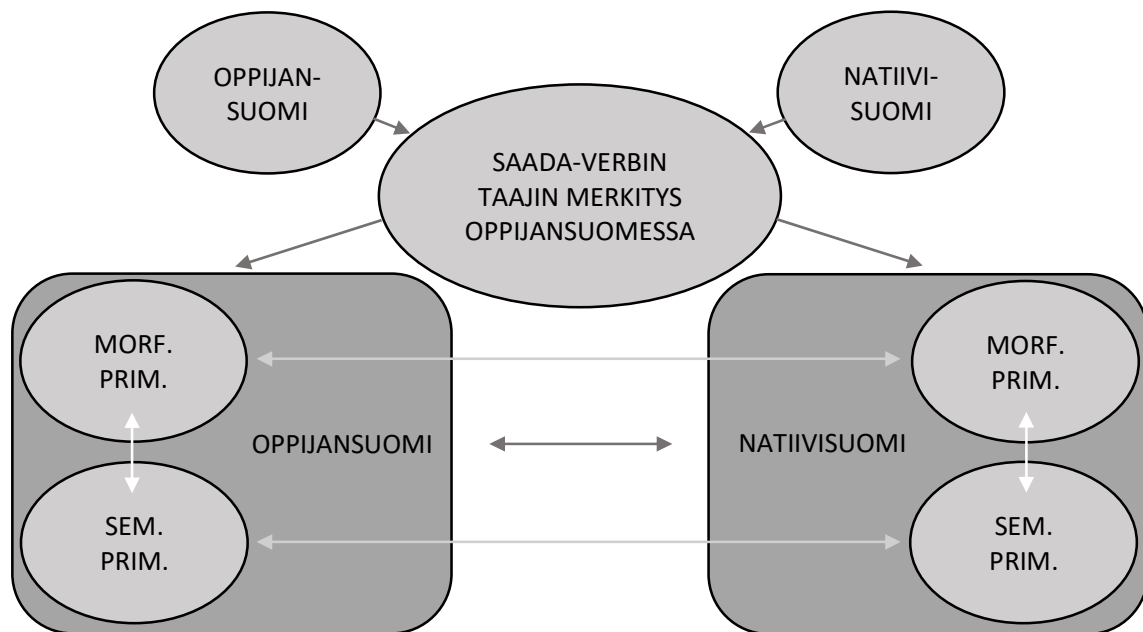
Globaalissa maailmassa kuulee ja näkee yhä useammin suomea, jonka on tuottanut henkilö, jonka äidinkieli ei ole suomi. Suomessa on kasvava määrä maahanmuuttajia, jotka oppivat suomea toisena kielenä, mutta lisäksi suomen kieltä opiskellaan maailmalla vieraana kielenä: suomea voi opiskella yli 100 yliopistossa noin 30 maassa (Opetushallitus 2018). Tutkimuksen painopiste on juuri suomi vieraana kielenä -opiskelijoiden tuottamassa oppijansuomessa, jolloin opiskelijat ovat oppineet suomea vieraskielisessä ympäristössä kieliyhteisön ulkopuolella.

Tässä maisterintutkielmassa verrataan suomi vieraana kielenä -aineistoa suomea äidinkielenään puhuvien tuottamaan aineistoon pyrkimyksenä selvittää, kuinka kielimuodot eroavat toisinaan. Jatkossa suomi vieraana kielenä -aineistosta käytetään nimitystä oppijansuomen aineisto ja äidinkielisestä aineistosta nimitystä natiivisuomen aineisto. Oppijansuomen aineisto on kerätty Kansainvälisestä oppijansuomen korpuksesta (ICLFI) (Jantunen 2011b) ja natiivisuomen aineisto Käännössuomen korpuksen (KSK) alkusuomen aineistosta (Jantunen 2004: 5 mukaan Mauranen 2000). Tutkimus sijoittuu kontekstuaalisen semantiikan viitekehykseen, tarkemmin fraseologiaan. Kontekstuaalisessa semantiikassa nimensä mukaisesti korostetaan kontekstin merkitystä tutkimuksessa. Fraseologiassa pyritään paljastamaan autenttisia kielessä piileviä rakenteita. Vertailumenetelmänä käytetään kontrastiivista oppijankielen tutkimusta (CIA) (ks. Granger 2015) ja käytännössä tutkimus toteutetaan korpuspohjaisella tutkimusmenetelmällä (ks. Jantunen 2011a; McEnery & Hardie 2014), jossa sekä kvantitatiivinen että kvalitatiivinen analyysi yhdistyvät. Tilastollisina menetelmänä hyödynnetään z-testiä X^2 -riippumattomuustestiä.

Tutkimus risteää myös oppijankielen tutkimuksen kanssa. Fraseologisella oppijankielen korpustutkimuksella on vakiintunut asema englannin kielen tutkimuksessa (ks. Granger 1998; Nesselhauf 2003; 2005; Meunier & Granger 2008a), ja se on saanut jalansijaa myös suomen kielen tutkimuksessa 2000-luvulla (ks. Jantunen 2009b, Jantunen & Brunni 2012). Aiheesta tehdään myös yhä enemmän maisterintutkielma (ks. Seppälä 2012; Varrio 2014; Kuuluvainen 2015) ja väitöskirjoja (ks. Ivaska 2015). Tämä maisterintutkielma sijoittuu tähän jatkumoon.

Oppijan- ja natiivikielen eroja tarkastellaan oppijansuomen aineiston yleisimmän ja natiivisuomen aineiston vastaavan SAADA-verbin merkityksen fraseologisten ominaisuuksien, morfologisen ja semanttisen primingin, suhteen. Morfologista primingia on paradigmaattista ja

kontekstuaalista. Paradigmaattisella morfologisella primingilla tarkoitetaan tässä tutkimuksessa ilmiötä, jossa lekseemi (tässä tapauksessa SAADA) suosii tiettyä sananmuotoa (ks. luku 2.2.4). Kontekstuaalinen morfologinen priming sen sijaan tarkoittaa lemman sananmuotojen suhdetta kontekstiin. Semanttisella primingilla puolestaan tarkoitetaan merkitysympäristöä, jonka kanssa lekseemi myötäesiintyy (ks. luku 2.2.5). Tutkimuksessa selvitetään myös, onko muotoa ja merkitystä edustavien fraseologisten yksiköiden välillä korrelaatiota¹ ja eroavatko mahdolliset löydetyt yhteydet oppijansuomen ja natiivisuomen välillä (ks. tutkimusasetelmasta kuvio 1).



Kuvio 1. Tutkimusasetelma.

Tutkimustuloksia oppijan- ja natiivikielen eroista voi hyödyntää suomi toisena ja vieraana -kielenä opetuksessa. Tutkimuksessa ei kuitenkaan käsitellä kielen oppimisprosessia, vaan valmiita oppijoiden tuotoksia. Tutkimustulosten soveltamisessa on kuitenkin pedagoginen intressi. Etenkin, mikäli kielimuotojen väliltä löytyy systemaattisia eroja, niiden huomioonottaminen opetuksessa voi auttaa oppijaa pääsemään lähemmäs natiivinkaltaista kielitaitoa. Natiivinkaltaista kielitaitoa voi tosin pitää illuusiona, sillä ”kieltä ei ole olemassa yhtenäisenä systeeminä” (Mustonen 2015: 75). Lisäksi voi kyseenalaistaa, onko natiivinkaltainen kielitaito edes tavoiteltavaa. Kielenoppimisella ei ole päätepistettä, vaan oppiminen on jatkuva prosessi, eivätkä edes natiivit puhujat saavuta kielen ”ylintä abstraktitasoa”. (Mp. 53.)

¹ Tässä tutkimuksessa korrelaatiolla tarkoitetaan kahden muuttujan välistä tilastollista riippuvuutta (ks. Heikkilä 2014: 12), ei kielitieteellistä korrelaatiota (ks. VISK § 106).

Tutkimuksen tavoitteena on siis selvittää, kuinka SAADA-verbin käyttö eroaa fraseologisesti oppijansuomessa ja natiivisuomessa. Tutkimus toteutetaan vertailemalla oppijankielen ja natiivikielen SAADA-verbin morfologista ja semanttista primingia sekä näiden fraseologisten yksiköiden välistä korrelaatiota seuraavien tutkimuskysymysten avulla:

1. Missä merkityksissä SAADA-verbi esiintyy taajimmin
 - a) oppijankielessä b) natiivikielessä?
2. Millaista paradigmaattista morfologista primingia oppijankielen SAADA-verbin taajimmalla ja natiivikielen vastaavalla merkityksellä esiintyy
 - a) oppijankielessä b) natiivikielessä?
3. Millaista semanttista primingia oppijankielen SAADA-verbin taajimmalla ja natiivikielen vastaavalla merkityksellä esiintyy
 - a) oppijankielessä b) natiivikielessä?
4. Onko paradigmaattisen morfologisen primingin ja semanttisen primingin välillä korrelaatiota?
 - a) oppijankielessä b) natiivikielessä?

Ensimmäisessä tutkimuskysymyksessä on kyse polysemiasta, ja sillä selvitetään SAADA-verbin eri merkitysten taajuudet oppijan- ja natiivisuomen aineistoissa. Polysemiassa on kyse sanaston sisäisestä merkityssuhteesta, jossa samalla sanalla on ”useita enemmän tai vähemmän toisistaan eroavia merkityksiä” (Kangasniemi 1997: 44) – tai kuten Cruse (1986: 23) asian ilmaisee, sanalla on ”jakautunut persoonallisuus”. Ensimmäinen tutkimuskysymys rajaa tutkimuksen koskemaan vain SAADA-verbin yleisintä oppijansuomen ja natiivisuomen vastaavaa merkitystä. Loput tutkimuskysymykset rakentuvat tämän asetelman varaan. Tutkimuksen edetessä sanan sisäisestä merkityksestä käytetään termiä merkitysluokka.

Toisen kysymyksen funktio on selvittää ensimmäisessä kysymyksessä selvinneen SAADA-verbin taajimman merkityksen paradigmaattista morfologista primingia. Tämä kysymys edustaa muodon ulottuvuutta muodon ja merkityksen välisessä suhteessa. Kolmannen kysymyksen tarkoitus on selvittää SAADA-verbin taajimman merkitysluokan semanttista primingia. Tämä kysymys edustaa merkityksen ulottuvuutta muodon ja merkityksen välisessä suhteessa. Semanttisesta primingista eli kontekstin merkityspiirteistä puhuttaessa käytetään termiä semanttinen luokka (vrt. polyseemisyyteen liittyvä termi merkitysluokka).

Viimeisen tutkimuskysymyksen tarkoitus on syventää ymmärrystä kaikista aiemmista tutkimuskysymyksistä, eli selvittää tilastollisin menetelmin muodon (morfologinen priming) ja merkityksen (morfologinen priming) välistä suhdetta SAADA-verbin oppijansuomen aineiston taajimmasta merkityksestä. Tämä tutkimuskysymys liittyy kontekstuaaliseen morfologiseen primingiin, mutta sillä myös haastetaan vallitseva fraseologisen yksikön kuvaus.

Tutkimuksessa on osin uudenlainen ja myös kokeileva lähestymistapa fraseologiseen tutkimukseen: verbin polyseemisuutta ja morfologista primingia on tarkasteltu ennenkin yhdessä (Jantunen & Brunni 2012), mutta tässä tutkimuksessa yhtälöön on lisätty vielä eksplisiittinen

kontekstin semanttinen tarkastelu ja riippuvuudesta fraseologisten elementtien välille yhdestä tietyistä SAADA-verbin merkityksestä. Ennalta tiedetään, että eri sananmuodoilla on erilaista semanttista primingia ja että verbien polyseemisessä käytössä on eroja oppijan- ja natiivisuomen välillä, mikä on odotuksenmukaista tässäkin tutkimuksessa (ks. Hoey 2005; Jantunen & Brunni 2012). Pelkkä havaintoyksiköiden tarkastelu ei kuitenkaan riitä selittämään morfologista vaihtelua, vaan selitys löytyy kontekstin ja havaintoyksikön suhteesta (Arppe 2002: 31), minkä vuoksi morfologista vaihtelua tarkastellaan sekä sanan sisäisen että kontekstuaalisen merkitysten kanssa yhdessä. Nyt huomio kiinnittyy myös sananmuotojen ja merkitysympäristön väliseen tilastolliseen yhteyteen, mitä ei ole ennen tehty.

1.2 SAADA-verbi tutkimuskohteena

Jotta tutkimuksella olisi mahdollisimman suuri merkitys suomi toisena ja vieraana -kielenä opetuksessa, tarkasteltavan lingvistisen elementin tulisi olla mahdollisimman taajaan käytetty. Taajuuden lisäksi verbin valitseminen puoltaa paikkansa siksi, että verbi on lauseen ydin ja määrittää siten kaiken muun lauseessa tapahtuvan (VISK § 444). Näiden seikkojen vuoksi tutkimukseen valikoitui SAADA-verbi, joka on Suomen sanomalehtikielen taajuussanaston (Kielipankki 2004) mukaan yleisin verbi heti verbien OLLA- ja EI-verbien jälkeen. Kun kaikki sanat otetaan huomioon, SAADA on sijalla 10/9996 (mp.). Kyseinen taajuussanasto koskee vain sanomalehtikieltä, mutta sen perusteella voi kuitenkin olettaa SAADA-verbin olevan taajaan käytetty myös muissa kirjoitetun kielen genressä.

Kuten taajaan käytetyt verbit hyvin usein ovat, myös SAADA on hyvin polyseeminen. Polyseemisen luokittelun pohjana käytetään Kielitoimiston sanakirjan (KS s.v. *saada*) määrittelemää 15:tä eri merkitystä (ks. tarkemmin luku 6.1.1 ja liite 1):

1. tapahtumasta jossa jk esine, hyödyke, oikeus tms. siirtyy jkn t. jnk omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jk tulee.
 - a. jnk lähetetyn vastaanottamisesta.
 - b. jnk oikeuden, mahdollisuuden, etuisuuden tms. tulemisesta jkn osaksi.
2. tulla, joutua jnk toimenpiteen, käsittelyn, kohtelun tms. kohteeksi.
 - a. joutua kurituksen, pahoinpitelyn, moitteiden tms. kohteeksi.
 - b. tulla tuomitukseksi rangaistukseen.
 - c. tulla, joutua kiittävän, moittivan tms. suhtautumisen kohteeksi.
 - d. (koulu)arvosteluun t. arviointiin liittyen.
3. ilmaantua t. tulla jnk osaksi t. hyväksi.
 - a. jälkeläisen syntymästä.
 - b. sairauteen, oireeseen t. vammaan liittyen.
 - c. *ark.* hämmästyttä, raivostumista, pelästymistä, inhoamista tms. merkitsevissä ilmauksissa.
 - d. muuttua, kehittyä jonkinlaiseksi.
4. onnistua hankkimaan jtkak itselleen, käyttöönsä, yhteyteensä tms., tavoittaa, saavuttaa.

- a. onnistua elatuksen, sadon, saaliin tms. hankkimisessa.
 - b. onnistua hankkimaan jku seurakseen t. ystävyys- tm. suhteeseen.
 - c. onnistua hankkimaan jku puolisokseen.
 - d. *ark.* päästä (naisen kanssa) yhdyntään.
 - e. *ark.* orgasmin saavuttamisesta.
5. jnk tulemisesta jkn aistimusten, tiedon t. ymmärryksen piiriin.
 6. tuloksen syntymisestä laskemalla, päättelemällä tms. tavalla.
 7. jnk aseman, olotilan, tehtävän tms. tulemisesta jnk osaksi; saavuttaa, tavoittaa.
 8. onnistua, kyetä tekemään t. aiheuttamaan jtk t. saattamaan tekeminen päätökseen.
 9. ilmaisemassa että jk on luvallista, sallittua. *Erik.* kohteliaisuutta ilmaisemassa.
 10. ilmaisemassa, että jk sopii, ei tuota haittaa, on yhdentekevää.
 11. ilmaisemassa tilaisuutta, mahdollisuutta jhk; voida, saattaa.
 12. ilmaisemassa jnk pakollisuutta t. väistämättömyyttä; joutua tekemään jtk.
 13. ilmaisemassa (moitteeseen liittyvää) käskyä, kehotusta t. toivomusta.
 14. nähdä-verbin ohella eräissä tulevaisuuteen viittaavissa ilmauksissa. *Erik.* välttelevissä vastauksissa.
 15. *vanh., runok.*
 - a. tulla, saapua.
 - b. ryhtyä, ruveta.

SAADA-verbiä ei ole aiemmin tutkittu kovin kattavasti, mutta siitä löytyy muutamia tutkimuksia. Esimerkiksi suomen kielen opetuksen ja oppimisen tueksi kehitetty verkkosanakirja ConLexis (ks. Jantunen, Kumpulainen, Tammimies, Tokola 2013) esittää SAADA-verbille kollokaateiksi lekseemit AIKA, RAHA, SURMA, TIETO, TUKI, VALMIS ja YRITTÄÄ. Tämä tarkoittaa sitä, että SAADA-verbi esiintyy usein edellä mainittujen sanojen yhteydessä. Lisäksi SAADA-verbistä on selvitetty, että sen partitiiviobjektien semanttiset roolit näyttäisivät suosivan 'vastaanottavuutta' ja 'tuloksellisuutta' (Tarvainen 2015).

SAADA-verbi sisältyy myös Pajunen (2001) argumenttirakennekuvaukseen. Pajunen (mts. 94, 264) luokittelee SAADA-verbin primääri-A-verbiksi, eli konkreettista asiantilaa kuvaavaksi verbiksi. Primääri-A-verbis kielentävät konkreettisia asiointiloja, mutta merkitykset voivat myös abstraktistua, jolloin verbis leksikalisoivat sekä konkreettisia että abstrakteja asiointiloja (mts. 94) – esimerkiksi juuri SAADA-verbi on laajentunut konkreettisesta merkityksestään myös abstrakteihin merkityksiin. Koska kyseessä on ensisijaisesti konkreettinen kuvaus, primääri-A-verbien osallistuja-argumentit eli valenssiin kuuluvat pakolliset täydennykset (ks. VISK § 446) viittaavat ensisijaisesti olioihin (Pajunen 2001: 94.) Tarkemmin SAADA kuuluu liikeverbeihin, koska omistussuhteen muutos on abstraktia liikettä (mp.). Liikeverbien sisällä SAADA jakautuu edelleen antamis-, ottamis- ja saamisverbeihin (mts. 264).

Antamisverbis kielentävät yleensä kädellä tehtävää tekoa, jonka seurauksena esine vaihtaa omistajaa joko vapaaehtoisuuteen tai pakkoon perustuen (Pajunen 2001: 262–263). Antamisverbis ovat kolmipaikkaisia: niillä on kaksi henkilötarkotteista argumenttia (x ja z) ja kolmantena antamisen kohdetta ilmaiseva argumentti (y). SAADA-verbillä x-argumentti on

subjekti, y-argumentti objekti ja z-argumentti erosijainen obliikvi² (mts. 264). SAADA-verbillä x-argumentti on resipientti eli vastaanottaja ja kohde usein jokin esine. SAADA-verbillä x, y- ja z-argumentit ovat aina edellytetyjä, mutta niiden kuvaama tilanne voidaan kuitenkin kielen-
tää siten, että esim. y- tai z-argumentti jää eksplisiittisesti mainitsematta. (Mts. 262–263.)
SAADA-verbin argumentit ovat siis yksinkertaistettuna vastaanottaja (x), kohde (y) sekä lähde (z), jotka ilmentyvät esimerkkilauseessa *minä sain sinulta kirjeen* seuraavasti: x = *minä*, y = *kirjeen* ja z = *sinulta*.

Pajusen työssä verbin polysemiaan ei kiinnitetä huomiota, koska usein polysemian lähtökohta on verbin konkreettisuudessa merkityksessä eikä argumenttirakenteen kuvauksessa ole mielekäästä esittää verbistä kaikkea tietoa (Pajunen 2001: 19). Tässä tutkimuksessa argumenttirakenteen kuvaus ilman polysemian huomioon otamista nähdään kuitenkin jossain määrin ongelmallisena, koska valenssi perustuu verbin merkitykseen ja koska verbit ovat usein monimerkityksisiä, samalla verbin eri merkityksillä voi olla erilainen täydennysympäristö (VISK § 449). Tästäkin tutkimuksessa polysemia on merkittävässä roolissa, mutta koska oppijansuomen taajin merkitys on juuri SAADA-verbin konkreettinen merkitys, sopii Pajusen kuvaus tämän tutkimuksen tarpeisiin.

Tutkielma etenee siten, että seuraavaksi tarkastellaan teoreettista viitekehystä (luvut 2 ja 3). Luvussa 2 esitellään kontekstuaalisen semantiikan sekä fraseologian lähtökohtia ja käsitellään tutkimuksen kannalta relevanttia terminologiaa. Luvussa 3 käsitellään oppijankielen määritelmää ja piirteitä sekä tarkennetaan fraseologista näkökulmaa oppijansuomen kontekstiin. Tämän jälkeen esitellään tutkimuksen aineistot Kansainvälinen oppijansuomen korpus sekä Käännössuomen korpus (luku 4). Aineiston jälkeen perehdytään tutkimusmetodeihin, kontrasttiiviseen oppijansuomen analyysiin, korpuspohjaiseen tutkimukseen ja tilastollisiin menetelmiin (luku 5). Luvussa 6 analysoidaan tutkimustuloksia eli vastataan aiemmin esitettyihin tutkimuskysymyksiin. Lopuksi (luvussa 7) pohditaan tuloksia ja niiden sovellutusmahdollisuuksia, arvioidaan tutkimuksen onnistumista sekä hahmotellaan aiheesta tarvittavaa jatkotutkimusta.

² Obliikvilauseenjäsenenellä tarkoitetaan verbin valenssiin kuuluvia argumentteja, jotka eivät ole subjekteja ja objekteja (Ojutkangas, Larjavaara, Miestamo & Ylikoski 2009: 139).

2 FRASEOLOGINEN VIITEKEHYS

2.1 Kontekstuaalinen semantiikka

Vallitsevan käsityksen mukaan kielijärjestelmän rakenne on jaettu karkeasti leksikkoon ja kielioppiin (ks. Hoey 2005: 1; Karlsson 2006 [1994]: 21; Kuiri 2012: 8). Perinteisesti on ajateltu, että leksikon eli sanaston sisältö on periaatteessa mahdollista luetella ja kielioppi sääntelee sitä, kuinka lekseemejä eli sanoja tulisi käyttää kielessä. Näin ei kuitenkaan ajatella tässä tutkimuksessa. Kantavana ajatuksena on Hoeyn (2005) näkemys siitä, että kielioppi syntyykin sanojen käytösäännöistä. Kaikki tämä on kuitenkin semantiikkaa: sanat merkityksineen sekä syntaksi ja morfologia, sillä sanojen yhdistäminen toisiinsa muuttaa ympäröivien sanojen merkityksiä ja näin luo isompia merkityskokonaisuuksia (Kuiri 2012: 9). Semantiikka on laaja kielen tarkastelun taso, joka jakautuu näkökulmasta riippuen myös useaksi alalajiksi. Näitä ovat muun muassa tässä tutkimuksessa hyödynnetty kontekstuaalinen semantiikka, mutta myös esimerkiksi lausesemantiikka, teoreettinen ja deskriptiivinen semantiikka ja generatiivinen ja kognitiivinen semantiikka (Kangasniemi 1997: 22–23).

Kontekstuaalinen semantiikka pyrkii nimensä mukaisesti selvittämään merkityksen muodostumista kontekstilähtöisesti (Jantunen 2004: 7). Kontekstuaalisen semantiikan periaatteet ohjaavat tätä tutkimusta, joka toteutetaan tarkemmin fraseologisen korpustutkimuksen viitekehysessä. Fraseologia on kielitieteen ala, jossa tutkitaan kielen vakiintuneita rakenteita, kuten monisanaisia leksikaalisia yksiköitä (ks. Granger 1998; Jantunen 2009b: 359–360).

Tässä tutkimuksessa ydinkäsitteitä ovat morfologinen priming (ks. luku 2.2.4) ja semanttinen priming (ks. luku 2.2.5), jotka ovat erityyppisiä fraseologisia yksiköitä ja jotka molemmat ilmentävät tutkimuksen kantavana ajatuksena olevaa primingin (ks. luku 2.2.3) teoriaa. Morfologisen ja semanttisen primingin avulla pyritään selvittämään SAADA-verbiin liittyviä vakiintuneita käytänteitä kontekstilähtöisesti, jolloin kyse on kontekstuaalisesta semantiikasta. Kontekstilähtöinen tarkastelu on perusteltua, koska sanat saavat lopullisen merkityksensä vasta kontekstissaan (esim. Ullmann 1967: 64; Kangasniemi 1997: 20; Cruse 2000: 14) – tai kuten Firthin (1968 [1957]: 179) kuuluisa ilmaus sanojen välisistä suhteista kuuluu: ”You shall know a word by the company it keeps!”

2.1.1 Lähtökohdat ja periaatteet

Kontekstuaalisen semantiikan tavoitteena on selvittää merkityksen muodostumista kontekstilähtöisesti. Konteksti voi olla ilmausta lähin tekstikonteksti tai laajempi, sosiaalisen ympäristön huomioon ottava. (Jantunen 2004: 7.) Kontekstuaalisuuden lähikäsitteellä kontekstilla sen sijaan tarkoitetaan lauseyhteyttä, eli lähintä tekstikontekstia. Tässä tutkimuksessa analyysi keskittyy juuri kontekstiin, mutta tutkimustulosten arvioinnissa pyritään ottamaan huomioon myös laajempi konteksti. Merkityksen kontekstuaalisen luonteen huomioonottaminen tutkimuksessa on brittiläisen koulukunnan peruja (Stubbs 1996: 25; Jantunen 2004: 7), ja varhaisimpana suuntauksen edustajana voi pitää J. R. Firthiä (ks. 1957; 1968 [1957]) (Stubbs 1996: 22, 25; Kenny 2001: 8).

Firth painottaa, että kielentutkimuksen pitäisi perustua autenttisiin kielenkäyttötilanteisiin ja kieltä pitäisi myös pyrkiä kuvaamaan tilannesidonnaisten relaatioiden avulla, koska tutkijan ei ole mahdollista päästä kiinni kielen taustalla vaikuttaviin prosesseihin tarkimmillakaan introspektiivisillä havainnoilla. Lisäksi kontekstuaalinen lähestymistapa mahdollistaa yleistämisen, kun analyysi ei nojaudu yksittäisen kielenkäyttäjän havaintoihin. (Firth 1968 [1957]: 170, 176.) Firthin mukaan kieli on siis sosiaalinen ihmisen toiminnan muoto ja kieli kuuluu aina kielenkäyttötilanteeseen³, joka määrittelee ilmausten merkityksen. Firthin käyttämä abstrakti käsite tilannekonteksti (*context of situation*) tarkoittaa sosiaalisten prosessien osanottajia, heidän verbaalista ja nonverbaalista toimintaansa, muita tilanteen olennaisia piirteitä sekä verbaalisen toiminnan vaikutuksia. Kaikki mainitut neljä tekijää ovat yhteydessä toisiinsa ja niiden toistuessa kielessä usein tietyllä tavalla ne muodostavat tilannekontekstin. (Firth 1957: 182; 1968 [1957]: 175–176.) Firthin lisäksi Stubbs (1996: 23) huomauttaa, että merkitystä ei pidä tutkia kontekstista irrotetuista tekstikatkelmista, vaan konteksti on aina otettava huomioon.

Firth (1957: 35; 1968 [1957]: 170) myös näkee, että tutkittavan kielen täytyy olla objektiivisesti havainnoitavissa, koska merkitykset eivät liity piilossa oleviin mentaaliin prosesseihin, vaan tilannekontekstiin ja havainnoitavaan kieleen (vrt. Hoey 2005). Tämän pohjalta Jantunen (2004: 7–9) huomauttaa, että kontekstuaalinen semantiikka eroaa erityisesti konseptuaalisesta semantiikasta. Konseptuaalisen semantiikan mukaan merkitys on käsite, idea tai mielikuva, joka syntyy kielenkäyttäjän tajunnassa sanasta ja jonka kielenkäyttäjä yhdistää merkitykseen (Kangasniemi 1997: 10). Kontekstuaalinen ja konseptuaalinen tarkastelutapa eivät sulje toisiaan pois, vaan voivat jopa täydentää toisiaan, kuten Hoey (2005) leksikaalisen primingin

³ Alun perin Malinowski käytti tilannekontekstin käsitettä hyväkseen antropologisessa tutkimuksessa, ja Firth on lainannut käsitteen häneltä (Firth 1957: 181).

teoriassaan osoittaa. Toisaalta korpuslingvistinen kontekstuaalinen merkityksen tutkimuskin tarvitsee käsiteanalyysia tuekseen, etenkin semanttista primingia analysoitaessa (Jantunen 2004: 9–10). Tämän tutkimuksen analyysin tarkastelukohteina ovat muun muassa verbin polysemia ja semanttinen priming, minkä vuoksi tutkimuksessa on konseptuaalisia piirteitä kontekstuaalisuuden lisäksi.

Firthin kanssa objektiivisesta lingvistiksestä tutkimuksesta samoilla linjoilla on Sinclair (1991), jonka mukaan introspektioon perustuvat havainnot eivät riitä tarjoamaan tietoa kielenkäytöstä. Hänen mukaansa esimerkiksi sanakirjan teossa ei voi käyttää lähteenä ainoastaan kielenkäyttäjien ajatuksia kielestä, sillä tällä tavoin kerättäisiin käsityksiä faktojen asemesta. Sinclair ei kuitenkaan kiellä tutkijan intuition tarpeellisuutta, vaan painottaa, että uskottava tutkimus ei voi pohjautua sille. (1991: 36–40.)

Kontekstuaaliselle semantiikalle on kielen autenttisen tarkastelun ja kontekstin huomioimisen lisäksi tyypillistä korostaa merkityksen, muodon, leksikon sekä kieliopin välistä yhteyttä (Jantunen 2004: 7). Sanan merkitys ja erilaiset kieliopilliset valinnat ovat sidoksissa toisiinsa, ja merkitys on puolestaan vahvasti sidoksissa rakenteeseen. Kieliopin ja leksikon kuvaukset täydentävät toisiaan, ja leksikaalisten yksiköiden systemaattinen myötäesiintymisen kuvaus mahdollistaa kielen sääntöjen ja rakenteiden kuvausten täydentämisen. Tämän vuoksi kielestä tulisikin etsiä säännöllisiä ja tyypillisiä yhteyksiä kieliopin ja leksikon väliltä. Kieliopin kuvauksista on kuitenkin varsin usein jätetty sanasto merkityksineen huomiotta. Sanakirjoissakin vain harvoin kiinnitetään huomiota muotoseikkoihin, kuten siihen, että sanan tietty muoto voi olla tyypillinen vain tietylle sanan merkitykselle. (Sinclair 1991: 6–7, 103–105.)

Sinclairin kanssa samaa mieltä on Hoey (2005: 1), jonka mukaan leksikko on rakenteeltaan kompleksinen ja systemaattinen, ja kielioppi onkin tämän leksikaalisen rakenteen tulos – eikä niin kuin perinteisesti on ajateltu, että kielioppi olisi ensin ja sanat ikään kuin pudoteltaisiin kieliopillisiin laatikoihin. Myös Stubbs (1996: 23) esittää, ettei kielioppia ja leksikkoa tulisi erottaa, vaan niitä tulee käsitellä toisistaan riippuvaisina – kuten ei muotoa ja merkitystäkään ei ole mielekäästi erottaa toisistaan.

Kontekstuaalisen semantiikan keskeinen ajatus on siis merkitystä tutkittaessa tarkastella autenttista, todenmukaista kielenkäyttöä (Stubbs 1996: 22–23), eli se pyrkii tutkimaan konkreettisia ja toistuvia käyttöyhteyksiä ja kuvaamaan niiden yleisiä merkityksiä. Tässä korpuslingvistiksestä tutkimuksesta sosiaalisen ympäristön huomioonottava analyysi ei ole mahdollista, mutta virkkeen laajuinen (ja tarvittaessa laajempi) konteksti on huomioitu kummassakin aineistossa. Korpuslingvistiksestä tutkimuksesta merkitystä pyritäänkin yleensä analysoimaan firtiläistä tilannekontekstia suppeamman kontekstin avulla (Jantunen 2004: 11), joten

toimintatapa ei ole poikkeuksellinen. Huomautettakoon Pirkolan (2016: 3) tapaan, että tilannekontekstin käsitteen keskeinen ajatus on, että tietyt asiat toistuvat kielessä ja on siis olemassa tyypillisiä tapoja käyttää kieltä. Tämä ajatus on tämän tutkimuksen ytimessä, kun pyrkimys on selvittää oppijankielen ja natiivikielen tyypillisiä tapoja käyttää SAADA-verbiä sekä sitä, kuinka nämä tyypilliset tavat eroavat toisistaan. Vaikka tutkimuksessa ei voi täysin huomioida firtiläistä tilannekontekstia, sitä mukaillaan: Aineistot ovat autenttista kieltä, mutta tutkijan intuitio on välttämätöntä tutkimuksen tekemisen kannalta. Etenkin aineiston luokittelussa (SAADA-verbi eri merkitysluokkiin sekä kontekstin sanat semanttisen primingin luokkiin) sekä tutkimustulosten analysoinnissa intuitiolla on merkittävä rooli. Kontekstuaalisen semantiikan vaatimus muodon ja merkityksen välisestä suhteesta toteutuu lähestymällä tutkimusaihetta polyseemisesti, mutta myös tutkimalla muodon ja merkityksen välistä suhdetta tilastollisesti.

2.1.2 Kontekstuaalinen valinta ja rajoitukset

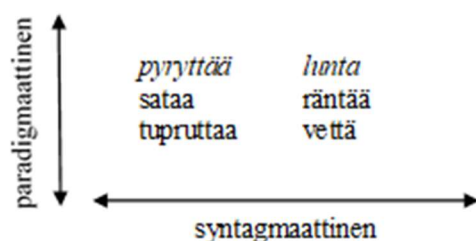
Kun kielenkäyttötilanteessa valikoidaan ilmausta, prosessia ohjaavat sekä kontekstuaaliset että sisäiset valinnat ja rajoitukset⁴, syntagmaattinen ja paradigmaattinen valinta. Syntagmaattinen ulottuvuus kuuluu kontekstuaalisiin valintoihin ja rajoituksiin (ks. Cruse 2000: 148–149). Kyse on lineaarisesta jatkumosta, johon jokin ilmaus valitaan peräkkäiseen ja jatkumolliseen suhteeseen muiden ilmausten kanssa (Jantunen 2004: 8). Perinteisesti on ajateltu, että syntagmaattisten suhteiden rakentumista säätelee syntaksi (ks. Tieteen termipankki s. v. *syntagmaattinen suhde*), jolloin syntagmaattisessa suhteessa olevat sanat ovat usein myös tiiviissä syntaktisessa suhteessa keskenään (ks. Cruse 2000: 148). Tässä tutkimuksessa nojataan kuitenkin Hoeyn (ks. 2005) leksikaalisen primingin teoriaan, jossa syntagmaattiset suhteet rakentuvat pikemminkin sanojen käytösäännöistä.

Sisäiset valinnat ja rajoitukset sisältyvät paradigmaattiseen ulottuvuuteen (ks. Cruse 2000: 148; Sinclair 2004: 140–146). Paradigmaattiseen ulottuvuuteen ilmaus valitaan käytettäväksi muiden mahdollisten ilmausten, kuten sanan kanssa paradigmaattiseen merkityssuhteeseen kuuluvien synonyymien ja hyponyymien, joukosta (Jantunen 2004: 8, 53). Lisäksi paradigmaattisessa suhteessa keskenään olevat sanat kuuluvat usein samaan syntaktiseen kategoriin (Cruse 2000: 148), ja paradigmaattinen valinta heijastelee lauseen tietyn position mahdollisia semanttisia valintoja (Cruse 2000: 148). Kuvion 2 esimerkissä paradigmaattisessa

⁴ Kontekstuaaliseen valintaan vaikuttaa myös lausekontekstia laajempi kehys, diskurssi: tyypilliset tavat keskustella (esimerkiksi akateemisissa diskurssissa) vaikuttavat kontekstuaaliseen valintaan (Stubbs 2001: 147). Tämän tutkimuksen puitteissa on kuitenkin mahdollista keskittyä vain kontekstin laajuiseen kontekstiin.

suhteessa toisiinsa olevat *lunta*, *räntää* ja *vettä* kaikki saneet ovat nomineja ja kuuluvat semanttisesti asioihin, joita voi 'sataa'.

Kontekstuaalisen semantiikan keskiössä on erityisesti tarkasteltavien ilmausten syntagmaattisten ja paradigmaattisten ulottuvuuksien (ks. kuvio 2) leikkauspiste, jossa kiinnostuksen kohde on niissä säännöissä ja preferensseissä, joiden perusteella ilmaus on valittu sekä syntagmaattiseen sekä paradigmaattiseen suhteeseen (Jantunen 2004: 8–9). Syntagmaattisten ja paradigmaattisten valinnan suhteen voi nähdä siten, että toimivat yhdessä: Crusen (2000: 149) mukaan syntagmaattiset suhteet rajoittavat paradigmaattisten suhteiden toimintatilaa, mutta uuden hoeylaisen ajattelutavan mukaisesti voisikin ajatella paradigmaattisten suhteiden ohjailevan syntagmaattisten suhteiden syntyä.



Kuvio 2. Syntagmaattinen ja paradigmaattinen ulottuvuus⁵.

Kuvion 2 lauseessa *Pyryttää lunta* lekseemit PYRYTTÄÄ ja LUMI ovat keskenään syntagmaattisessa, lineaarisessa suhteessa. Paradigmaattisessa suhteessa lekseimin PYRYTTÄÄ kanssa ovat esimerkiksi SATAA ja TUPRUTTAA, ja lekseimin LUNTA kanssa paradigmaattisessa suhteessa ovat muun muassa RÄNTÄ ja VETTÄ. Kontekstuaalisessa semantiikassa pyritään siis selvittämään, kuinka sanat yhdistyvät toistensa kanssa paradigmaattisiin ja syntagmaattisiin suhteisiin, ja miten ne valikoituvat konteksteihinsa.

Sinclair (mm. 1987, 1991, 1996) on kuvannut lekseemien kontekstuaalista valintaa vieläkin tarkemmin. Lekseemien yhdistymistä ohjaavat syntagmaattisten ja paradigmaattisten suhteiden lisäksi vapaan valinnan periaate (*open choice principle*) ja idiomaattisen valinnan periaate (*idiom principle*). Vapaan valinnan periaatteella tarkoitetaan sanojen mahdollisuutta yhdistyä vapaasti kieliopin tarkoitusten puitteissa, ja tätä kuvataankin usein aukkotäydennysmalliksi: tekstissä on aukkoja, joihin lisätään sopiva sana kontekstin rajoitteiden puitteissa. Vapaan valinnan periaate on normaali tapa kuvata kieltä, ja lähes kaikki kieliopit perustuvatkin siihen. (Sinclair 1991: 109–110). Sinclair (mts. 110) pitää vapaan valinnan mahdollisuutta kuitenkin lähinnä teoreettisena, sillä sanat eivät esiinny teksteissä satunnaisessa järjestyksessä. Vapaan

⁵ Kuvio on mukailtu Jantunen (2004: 9) mallista.

valinnan periaate ei myöskään tarjoa tarpeeksi todellisia rajoituksia peräkkäisille eli syntagmaattisille valinnoille. Tämän vuoksi hän pitää idiomaattista valintaa tekstien synnyn perustana.

Idiomaattisen valinnan periaate kattaa sellaiset rajoitukset, joihin vapaan valinnan periaate ei riitä vastaamaan: kielenkäyttäjillä on valmiina suuri määrä kielenulkoisesta todellisuudesta riippumattomia⁶ puolivalmiita konstruktioita, jotka ohjaavat kielenkäytön valintoja (Sinclair 1991: 109–110). Jantunen (2009b: 358) huomauttaa, että idiomaattista periaatetta ei pidä ymmärtää pelkästään idiomeihin liittyvänä, vaan se koskee myös yksittäisen sanan kontekstuaalisia suhteita. Vapaan valinnan periaate vastaa siis perinteistä ajattelua kielestä, jossa sanat pudotellaan kielioppisääntöjen mukaisesti paikalleen. Idiomaattisen valinnan periaate sen sijaan vastaa uudenlaista käsitystä kielestä, jossa kieliopin ajatellaan kumpuavan sanojen käyttösäännöistä eli leksikaalisesta primingista (ks. Hoey 2005).

Sanojen yhdistämistä tulisikin pitää jatkumona, jonka toisessa päässä on (hypoteettinen) täysin vapaa valinta ja toisessa täysin rajoittunut myötäesiintyminen. Näiden ääripäiden väliin jää alue, jossa vaihtelu on jossain määrin sallittua ja sanojen yhdistämistä ohjaavatkin rajoitusten (ja niiden puuttumisen) sijaan myötäesiintymispreferenssit. (Jantunen 2004: 15.) Jantunen (mp.) korostaakin, että rajoitukset, tendenssit ja preferenssit (priming) tulisi erottaa toisistaan, koska tyypillisimmin kyse on vahvasta preferenssistä (primingista) kuin ehdottomasta rajoituksesta.

Vapaan ja idiomaattisen valinnan käsitteet ovat tutkimuksessa oleellisia, koska pyrkimyksenä on selvittää SAADA-verbin semanttisia ja morfologisia valintapreferenssejä eli primingia, mutta myös välttelysuhteita. Morfologinen priming edustaa sekä paradigmaattista että syntagmaattista ulottuvuutta, kun taas semanttinen priming kuuluu syntagmaattiseen ulottuvuuteen – täytyy kuitenkin muistaa, että paradigmaattiset ja kontekstuaaliset valinnat ovat yhteydessä toisiinsa (Cruse 2000: 148). Tutkimuksen kannalta mielenkiintoista on nähdä, kuinka morfologinen ja semanttinen priming sekä välttelysuhteet eroavat oppijan- ja natiivisuomen välillä: toisaalta oppijalla ei ole intuitiivisesti hyödynnettävissä Sinclairin (1991: 110–111) mainitsemia puolivalmiita konstruktioita, mutta toisaalta kielenoppimisprosessin alussa ollaan hyvin vahvasti ulkoa opeteltujen fraasien varassa.

⁶ Sinclair (1991: 110) myöntää, että todellisuudessa yhdessä esiintyvät asiat esiintyvät todennäköisesti yhdessä myös kielessä, mutta korostaa, että kieli sisältää myös paljon todellisuudesta riippumattomia valintoja.

2.2 Fraseologia

2.2.1 Fraseologia tutkimusalana

Fraseologiassa tutkitaan vakiintuneita kielen rakenteita, joita ovat muun muassa idiomit, kollokaatiot ja rutiinit eli totutut ilmaukset (Granger & Paquot 2008: 27–28; Jantunen 2009b: 360–361), mutta myös muut vakiintuneet monisanaiset leksikaaliset yksiköt (ks. Granger 1998, 2005; Jantunen 2009b: 357). Fraseologia voidaan määritellä myös lemman tai lekseemin myötäesiintymisenä minkä tahansa lingvistisen elementin kanssa (näitä elementtejä voivat olla esimerkiksi toinen sananmuoto tai kieliopillinen kaava). Fraseologialle onkin tyypillistä tutkia kiellisten yksiköiden myötäesiintymistä. (Gries 2008: 4–5.) Näin ollen fraseologinen kielentutkimus täyttää kontekstuaalisen semantiikan periaatteet todellisen kielenkäytön ja kontekstuaalisten suhteiden tärkeydestä (Pirkola 2016: 7).

Usein fraseologia liitetään sanastontutkimuksen alaluokaksi, mutta siinä käsitellään pikemminkin sanaliittoja ja -yhdistelmiä kuin yksittäisiä sanoja (Granger & Meunier 2008b: XIX). Fraseologia on kuitenkin idiomeja, metaforisia ilmauksia ja sanontoja laajempi tutkimusalue, sillä fraseologia ilmenee natiivipuhujan kielenkäytössä laajoina sanastollis-kieliopillis-semanttisina ja tekstuaalisina ilmiöinä, esimerkiksi leksikaalisena primingina (*lexical priming*, ks. Hoey 2005) (Jantunen & Brunni 2012: 75).

Fraseologiaa voi luonnehtia myös rakenteen, merkityksen ja sanayhdistelmien käytön tutkimuksena (Cowie 1994: 3168). Toisaalta fraseologia leikkaa muiden alojen, kuten semantiikan, morfologian, syntaksin ja diskurssintutkimuksen kanssa, minkä vuoksi fraseologian rajat ovat häilyvät (Granger & Paquot 2008: 29): fraseologia on kytköksissä lähes kaikkeen, ja tämä seikka tekee fraseologiasta vaikean tutkimusalan, mutta siinä piilee myös sen vetovoima (Grangerin ja Paquotin 2008: 29 mukaan Mel'čuk 1995: 227). Edellä mainituista aloista fraseologia risteää ehdottomasti eniten semantiikan kanssa, mutta samaan aikaan raja fraseologian ja semantiikan välillä on epäselvä (Granger & Paquot 2008: 30). Eräs rajatapaus on tässäkin tutkimuksessa esiintyvä semanttinen priming. Sen sijaan toinen tässä tutkimuksessa tarkasteltu fraseologinen ilmiö, morfologinen priming, liittyy vahvasti morfologiaan. Tärkeää onkin huomata, että alat eivät ole omissa tyhjiöissään, vaan aina jollain tavalla limittyneitä muiden alojen kanssa.

Fraseologisessa tutkimuksessa on kaksi päähaaraa. Ensimmäinen on erityisesti itäeurooppalaisista maista ja Venäjältä kotoisin. Suuntauksessa on pyritty muun muassa tutkimaan

monisanaisten yksiköiden varioivuutta ja jähmeyttä. Tämän haaran perintönä voi pitää fraseologisen tutkimuksen vakiintumista omaksi tutkimushaarakseen sekä alan termistön syntyä (Granger & Paquot 2008: 28). Toinen suuntaus on tuoreempi, brittiläisen koulukunnan erityisesti korpusvetoinen tutkimus leksikaalisen myötäesiintymisen parissa. Suuntauksen isänä voi pitää Sinclairia (ks. esim. 1987), joka on tehnyt urauurtavaa tutkimusta leksikografian parissa ja uudistanut fraseologista näkökulmaa. (Cowie 1998: 1; Granger & Paquot 2008: 28–29.) Uudempi suuntaus on paljastanut paljon sanayhdistelmiä, jotka eivät mahdu aiemmin määriteltyihin kielellisiin kategorioihin, ja on siksi avannut paljon syntagmaattisia mahdollisuuksia tutkimuksen kannalta (mm. kollokaatit ja kolligaatiot) (Granger & Paquot 2008: 29). Ensimmäinen suuntaus pohjaa vahvasti kielitieteelliseen traditioon (*linguistically based*), kun taas jälkimmäinen on aineistovetoinen (*data-driven*) (Granger & Paquot 2008: 45).

Kiinnostus fraseologiseen tutkimukseen on herännyt muutaman viimeisen vuosikymmenen aikana, ja vasta tällä vuosituhanella fraseologia on vakiinnuttanut asemansa omana kielentutkimuksen alanaan (Granger & Meunier 2008b: XIX; Granger & Paquot 2008: 27; Gries 2008: 3), vaikka Firthin ajatukset kontekstuaalisesta semantiikasta sijoittuvatkin noin 60 vuoden taakse. Kiinnostusta fraseologiaan on erityisesti lisännyt helposti saatavilla olevat sähköiset tutkimusaineistot, korpuset. Ne ovat tärkeitä työkaluja fraseologiselle tutkimukselle, sillä ne mahdollistavat kvantitatiivisen lähestymistavan. (Sinclair 2008: XVI; Hoffmann, Fischer-Stracke & Sand 2013: 1). Korpustutkimus on oiva tapa lähestyä fraseologiaa, kielen vakiintuneita yksiköitä, koska korpustutkimuksella saa esille kielen piileviä (vakiintuneita) rakenteita (Jantunen 2009b: 373) ja sillä saa tietoa ilmiöiden esiintymisen ja myötäesiintymisen taajuuksista (Gries 2008: 15–16). Tässäkin tutkimuksessa fraseologiaa tutkitaan korpusaineistoin osin kvantitatiivisella lähestymistavalla.

Fraseologian luonnehdinta on parissa kymmenessä vuodessa muuttunut jähmeiden idiomien tutkimuksesta monimuotoisempaan suuntaan (Cowie 1994: 3168–3169; Gries 2008: 3). Nykyään fraseologian parissa tehdyllä tutkimuksella on suuri vaikutus tutkijoihin monella (kieli)tieteenhaaralla, kuten kielenoppimisen, -omaksumisen ja -opettamisen tutkimuksessa. Fraseologian ja sen tutkimustulosten vaikutusta ei kuitenkaan täysin tunnusteta ja tunnusteta tai kuvasteta terminologisesti näillä tutkimusaloilla (Gries 2008: 3). Griesin (mp.) mukaan tämä on valitettavaa, sillä edellä kuvatun tilanteen vuoksi ei ole helppoa tunnistaa alueita, joihin fraseologinen tutkimus on jättänyt jälkensä. Lisäksi tästä syystä eri aloilla tehdään limittäisiä oletuksia, konsepteja ja löydöksiä vähemmän läpinäkyviksi kuin olisi suotavaa. Väljän terminologian vuoksi on myös hyvin vaikea tunnistaa, milloin samankaltaisia asioita verrataan

toisiinsa (Wray & Perkins 2000: 3). Tämän vuoksi tässä tutkimuksessa pyritään avaamaan ja perustelemaan terminologiset valinnat tarkasti.

Fraseologiaa on siis vaikea määritellä yksiselitteisesti, koska eri tutkijat ovat antaneet sille erilaisia määritelmiä. Tämän lisäksi alan sisäinen terminologia on sekavaa, sillä samoille ilmiöille voi olla käytössä useita termejä ja sama termi voi viitata eri asioihin. (Granger & Paquot 2008: 27–28; Gries 2008: 4; Hoffmann ym. 2013: 1.) Tässä tutkimuksessa fraseologia nähdään piilevien vakiintuneiden ilmausten ja kielenkäyttötapojen tutkimisena muun muassa Grangerin (ks. 1998; 2005) ja Jantusen (ks. 2009b) tapaan ja kielellisten ilmiöiden myötäesiintymisen tutkimuksena Griesin (ks. 2008) ja Jantusen (ks. 2009b) tapaan. Nämä fraseologiset näkemykset ovat vahvasti linkittyneet myös Hoeyn (ks. 2005) teoriaan leksikaalisesta primingista, joka on tutkimuksen kantava näkemys kielestä. Terminologinen tavoite tällä tutkimuksella onkin kirjoittaa morfologinen ja semanttinen priming vahvemmin osaksi leksikaalisen primingin teoriaa.

2.2.2 Fraseologinen yksikkö

Fraseologinen yksikkö on yhtä sanaa laajempi elementti (Jantunen 2009b: 358) eli leksikaalis-kieliopillinen rakenne, joka sisältää lekseemin kontekstuaalisine valintoineen (Jantunen 2012: 205). Tällaisia kielellisten elementtien välisiä suhteita ovat esimerkiksi tässä tutkimuksessa tarkastellut morfologinen ja semanttinen priming. Kuitenkaan fraseologista yksikköä, kuten fraseologian tutkimualaakaan, ei määritellä yhdenmukaisesti. Fraseologinen tutkimus onkin keskittynyt paljon sellaisten kriteerien löytämiseen, joilla ylipäätään voitaisiin erotella tietyn tyyppinen fraseologinen yksikkö toisenlaisesta (Grangerin ja Paquotin 2008 mukaan Cowie 1998: 6). Jantunen (2004: 18–21; 2009b: 358–360) ja Gries (2008) ovat esittäneet näkemyksiään fraseologisen yksikön määrittelemiseksi. Griesin (2008: 4–10) esitys sisältää kuusi parametria, jotka ovat vapaasti suomennettuna:

1. fraseologiseen yksikköön sisältyvien elementtien luonne
2. fraseologiseen yksikköön sisältyvien elementtien määrä
3. montako kertaa ilmaus täytyy havaita, ennen kuin sen voi laskea fraseologiseksi yksiköksi
4. fraseologiseen yksikköön sisältyvien elementtien sallittu etäisyys
5. fraseologiseen yksikköön sisältyvien elementtien leksikaalisen ja syntaktisen vaihtelun määrä
6. semanttisen yhtenäisyyden ja epäkompositionalisuuden/ennustamattomuuden rooli määritelmässä

Seuraavaksi fraseologinen yksikkö määritellään siten, että käsittely rakentuu Griesin parametrien varaan, mutta tarkasteluun on otettu mukaan muidenkin näkemyksiä samasta aiheesta. Jokaisen parametrin käsittelyn lopuksi kerrotaan, kuinka kyseinen parametri näyttäytyy tässä tutkimuksessa.

1) fraseologiseen yksikköön sisältyvien elementtien luonne

Ensimmäisen parametrin mukaan on määriteltävä, millaisista elementeistä fraseologinen yksikkö voi muodostua. Griesin mukaan fraseologinen yksikkö on sananmuodon tai lemman ja jonkin muun kielellisen elementin myötäesiintymä – tämä toinen elementti voi olla toisen sanan lisäksi esimerkiksi jokin kieliopillinen rakenne. (Gries 2008: 5). Kahden sanan myötäesiintymästä Gries (mp.) antaa esimerkin *strong tea*. Kyse on kollokaatiosta, joka on yleensä tilastollisesti merkittävä kahden lekseemin muodostama leksikaalinen myötäesiintymä (Jantunen 2009b: 358). Jantunen (mp.) esittää kielenulkoisesta maailmasta motivoituneeksi esimerkiksi kollokaatiot KANA <KOTKOTTAA> ja KUKKO <KIEKUA>. Toisaalta kollokaatioilla ei kovin usein ole tällaista motivaatiota (mp.). Esimerkiksi adverbi KOVIN kollokoi EI-kieltosanan kanssa, mutta synonyymiset OIKEIN ja HYVIN eivät (Jantunen 2004: 86–92). Kollokaatiota on luonnehdittu myös psykologiseksi ilmiöksi, sillä luonnollista kollokaatiota ei esiintyisi, mikäli sanat olisivat erillään ihmisten mielissä (Hoey 2005: 7). (Ks. kollokaatiosta myös Palmer 1981: 75–79; Cowie 1994: 3169; Kenny 2001: 81.)

Sana voi myötäesiintyä toisten sanojen lisäksi myös kontekstissa esiintyvän kieliopillisen kategorian kanssa syntagmaattisessa suhteessa (Jantunen 2004: 22, 31; ks. myös Gries 2008: 5). Tätä kutsutaan kolligaatioksi (ks. Firth 1968 [1957]; Barnbrook 1996; Tognini-Bonelli 1996; Jantunen 2001; 2004). Hyvä esimerkki kolligaatiosta on, että astemääräite *kovin* esiintyy taajaan yhdessä kieltoverbin kanssa (Jantunen 2004: 147).

Griesin ensimmäisestä teesistä puuttuvat eksplisiittisinä mainintoina Jantunen (2004: 31; 2009b: 359) esittämät semanttinen prosodia ja semanttinen preferenssi (priming), sillä fraseologisena yksikkönä voi pitää myös sanan myötäesiintymistä kontekstuaalisten semanttisten piirteiden kanssa. Semanttinen prosodia on leksikaalis-semanttinen myötäesiintymä, ja sitä kutsutaan myös merkitysuraksi (Louw 1993: 157; Jantunen 2009b: 359–360). Semanttinen prosodia voi olla positiivinen, negatiivinen tai neutraali esiintymisympäristö (Jantunen 2009b: 359–360) tai Sinclairin (2004: 140–146) mukaan ilmauksen pragmaattinen merkitys, jonka vuoksi ilmaus on alun perin valittu käytettäväksi. Stubbsin (1995: 15–16) paljon lainattu esimerkki englannin kielen CAUSE-verbillä (suom. 'aiheuttaa') havainnollistaa, että sillä on negatiivinen semanttinen prosodia (käyttöympäristö). Tämä ilmenee verbin saamalla negatiivisilla kollokaateilla: CAUSE <DEATH, PROBLEM, TROUBLE>. Semanttinen preferenssi (priming) on semanttista prosodiaa tarkempi yhteys, leksikaalis-semanttinen myötäesiintymä (Jantunen 2009b: 359).

Tässä tutkimuksessa fraseologinen yksikkö määritellään Jantusen (2009b; 2012) mukaisesti yhtä sanaa laajemmaksi elementiksi (Jantunen 2009b: 358), joka sisältää lekseemin ja sen kontekstuaaliset valinnat (Jantunen 2012: 205) ja Griesin (2008: 5) tapaan sananmuodon tai lemman ja jonkin toisen kielellisen elementin myötäesiintymänä, mutta myös Hoeyn (2005: 13) mukaan näiden elementtien välttelysuhteena. Hoey (mp.) ei ulota välttelysuhdetta kaikkiin leksikaalisen primingin hypoteeseihinsa, kuten semanttisen assosiaation hypoteesiin. Tässä tutkimuksessa X^2 -riippumattomuustesti paljastaa kuitenkin morfologisen ja semanttisen primingin väliltä myötäesiintymissuhteiden lisäksi välttelysuhteita, minkä vuoksi välttelysuhde on otettu fraseologisen yksikön määritelmään mukaan. Tämän määritelmän valossa tässä tutkimuksessa toteutuu kolme erilaista myötäesiintymis- ja välttelysuhdetta yhden lemman (SAADA) kanssa. Ensimmäinen suhde on lemman ja sen sisäisen merkityksen (polysemia) välinen, toinen lemman ja sen morfologisen muodon (morfologinen priming) välinen ja kolmas lemman ja sen leksikaalis-semanttisen käyttöympäristön (semanttinen priming) välinen. Neljäs tutkimuksessa tarkasteltava suhde morfologisen ja semanttisen primingin välillä haastaa vallitsevan fraseologisen yksikön määritelmän, sillä kyseessä on kahden erilaisen kielellisen elementin välinen fraseologinen suhde ilman, että toinen elementti on lekseemi.

2) fraseologiseen yksikköön sisältyvien elementtien määrä

Griesin (2008: 4–5) toisen parametrin mukaan on päätettävä, voiko fraseologinen yksikkö koostua esimerkiksi vain kahdesta elementistä, vai useammasta. Griesin (2008: 5) mukaan fraseologinen yksikkö voi koostua useammasta kuin kahdesta elementistä (esimerkiksi *to eke out a living*, jossa Gries (mp.) huomauttaa olevan paikka verbille *to eke*, suoralle objektille *living* ja partikkelille *out*, joka täydentää transitiivista verbiä).

Jantunen (2004) käyttää ilmiöstä nimitystä klusteri (ks. käsitteestä myös Kenny 2001: 43–44). Klusterilla tarkoitetaan kahta sanaa pidempää sanaketjua (Jantunen 2004: 19; 2009b: 359). Myöhemmin Jantunen (2009b: 359) käyttää klustereista myös nimitystä n-grammi, jossa ilmaistaan klusterin sisältämien sanojen määrä. Esimerkiksi Jantunen (2004: 103) on havainnut, että KOVIN-sana esiintyy tyypillisesti ilmauksessa *ei olla kovin*, joka on 3-grammi. Maisterintutkielman laajuuden vuoksi tässä tutkimuksessa ei ole mahdollista selvittää, millaisissa klustereissa SAADA-verbi esiintyy, vaikka sekin olisi relevanttia.

Tässä tutkimuksessa fraseologinen yksikkö muodostuu periaatteessa aina kahden elementin välisestä suhteesta (lemma + sisäinen merkitys / morfologinen muoto / leksikaalis-semanttinen käyttöympäristö, morfologinen muoto + leksikaalis-semanttinen käyttöympäristö), mutta

käytännössä asia ei ole niin yksinkertainen. Ensimmäisessä tutkimuskysymyksessä läsnä on kaksi elementtiä: lemma ja lemman sisäinen merkitys. Koska loput tutkimuskysymykset pohjautuvat tarkasteltavan lemman yleisimpään sisäiseen merkitykseen, on tämä elementti läsnä jatkossakin. Toisessa tutkimuskysymyksessä, jossa tarkastellaan morfologista primingia, läsnä on siis jo kolme elementtiä: lemma + lemman sisäinen merkitys + morfologinen muoto. Samoin on kolmannessa tutkimuskysymyksessä, mutta morfologisen muodon tilalla on semanttinen priming. Neljännessä tutkimuskysymyksessä läsnä ovat nämä kaikki neljä elementtiä. Lisäksi semanttisen primingin analyysissa yhdeksi kokonaisuudeksi on tarpeen mukaan laskettu useampi sana (esimerkiksi etu- ja sukunimi -yhdistelmät ja kiinteät ilmaukset on käsitelty yhtenä kokonaisuutena, esim. *Vladimir Putin ja tänä päivänä*). Tässä tutkimuksessa fraseologiseen yksikköön voi siis sisältyä enemmän kuin kaksi elementtiä – tai pikemminkin tarkasteltavat elementit voivat itsestään koostua useammasta elementistä.

3) montako kertaa ilmaus täytyy havaita, ennen kuin sen voi laskea fraseologiseksi yksiköksi

Kolmanneksi Griesin (2008: 4–5) mukaan on määriteltävä, kuinka monesti ilmauksen on esiinnyttävä, jotta sen voisi laskea fraseologiseksi yksiköksi. Tässä kohtaa tulee muistaa yksi fraseologian määritelmistä, jonka mukaan fraseologiassa tutkitaan nimenomaan kielen vakiintuneita yksiköitä (ks. Granger 1998, 2005; Jantunen 2009b: 357): esimerkiksi kollokaation käsitteellä viitataan toistuviin sanojen yhteisesiintymiin (Jantunen 2004: 16), ja Firth (1968 [1957]: 180) puhuu tyypillisestä kollokaatiosta (*habitual collocation*). Tässä tutkimuksessa (muiden tutkimusten mukaan) puhutaan pikemminkin vahvoista ja voimakkaista preferensseistä ja primingeista kuin ehdottomista säännöistä (ks. Jantunen 2004: 138–140). Tämä ei kuitenkaan perustu tutkijan intuitioon, vaan tilastolliset testit auttavat tunnistamaan merkittäviä vakiintuneita yhteyksiä.

4) fraseologiseen yksikköön sisältyvien elementtien sallittu etäisyys

Neljänneksi on määriteltävä, kuinka kaukana toisistaan fraseologisen yksikön elementit voivat esiintyä: joskus niitä tutkitaan vain vierekkäisistä sanoista (Gries 2008: 5), kun taas sekä Jantunen (2004: 83) että Gries (2008: 5) hyväksyvät toisistaan kaukana olevat elementit (katkonaiset ilmaukset) myös fraseologisiksi yksiköiksi. Etäisyys riippuu myös siitä, millaisesta kontekstuaalisesta piirteestä on kyse. Esimerkiksi Jantunen (2004: 19) mukaan kollokaatiossa elementtien välissä voi olla muita sanoja, mutta klusterissa sanat ovat peräkkäin. Tutkimuskirjallisuudessa

ei ole kuitenkaan yhteisymmärrystä siitä, mikä on sopiva etäisyys lekseemien kontekstuaalisia valintoja tutkittaessa (Stubbs 1995: 8), ja optimaalinen tarkasteluväli vaihtelee eri tutkimuksissa (Jantunen 2004: 19).

Sanevälin lisäksi etäisyyttä voi rajata myös muilla keinoin, esimerkiksi kieliopilliset rakenteet huomioimalla. Tutkimuskirjallisuudessa yleinen trendi on kuitenkin jättää kieliopilliset rakenteet pois etäisyyden määrittelystä, sillä kieliopillinen rakenne vaatisi saneilta vierekkäisyyttä. Etäisyyttä voi rajata myös lause- ja virkerajojen mukaisesti. Usein alue rajataan virkkeen sisäiseksi, mutta poikkeuksiakin löytyy. (Jantunen 2004: 20.) Esimerkiksi Halliday ja Hasan (1976: 286) puhuvat kollokationaalisesta koheesiosta, joka on virkerajat ylittävää kollokointia. Ilmiö on esimerkki tekstien leksikaalisesta koheesiosta (ks. mts. 284–288).

Etäisyyden määrittely on tärkeää, sillä tarkastelualan rajaaminen vaikuttaa tutkimustuloksiin (Jantunen 2004: 19). Tässä tutkimuksessa SAADA-verbin esiintymistä on otettu huomioon virkkeen laajuinen konteksti, joiden puitteissa SAADA-verbit virkkeineen jaettiin eri merkitysryhmiin. Tarvittaessa oli mahdollista tarkistaa myös laajempi konteksti. Tästä eteenpäin tarkastelualue on rajattu verbin merkityksen mukaisesti. Paradigmaattisen morfologisen preferenssin analyysissä etäisyyttä ei ole rajattu, koska kyse on sanan sisäisestä ilmiöstä. Semanttisen primingin analyysissä tarkastelualue on 2:2 eli kaksi sanetta noodin vasemmalta ja oikealta puolelta, kuitenkin virkerajat huomioon ottaen. Tarkastelualue on melko suppea tutkielman laajuuden vuoksi. Morfologisen ja semanttisen primingin korrelaation tarkastelussa sananmuodon ja semanttisen luokan etäisyys on sananmuodosta kaksi sanetta molempiin suuntiin.

Toisenlaisessa tutkimuksessa semanttista primingia voisi tutkia laajemmastakin kontekstista esimerkiksi avainsana-analyysin avulla (ks. Jantunen 2018), mutta tässä tutkimuksessa jokainen tarkastelualan sana luokiteltiin erikseen (ks. Pirkola 2016), minkä vuoksi alue ei voinut olla kovin suuri. Position sidottu semanttisen primingin tarkastelu ei ole kuitenkaan poikkeavaa (ks. Jantunen 2004).

5) fraseologiseen yksikköön sisältyvien elementtien leksikaalisen ja syntaktisen vaihtelun määrä

Viidenneksi tulee päättää, kuinka paljon fraseologisen yksikön elementit voivat joustaa syntaktisesti ja leksikaalisesti. Griesin (mts. 5–6) mukaan esimerkiksi *kick the bucket* -ilmaus (suom. 'potkaista tyhjää') sallii ajallista muuntelua, kun taas ilmaus *by and large* (suom. 'yleisesti ottaen') ei joustaa lainkaan. Lisäksi on leksikaalisesti vapaita ja joustavia rakenteita, joista Gries

(mp.) antaa esimerkiksi englannin kielen distransiitiivisen rakenteen, jonka ainoa ehto on, että siihen kuuluu subjekti, verbi ja kaksi objektia – näihin kohtiin voi sijoittaa monia eri lekseemejä, mikä tekee rakenteesta hyvin joustavan.

Tässä tutkimuksessa fraseologisella yksiköllä on leksikaalinen rajoite, sillä kaikissa tutkimuksessa esiintyvissä kielellisissä elementeissä on SAADA-verbin jokin sananmuoto tai ne esiintyvät sen läheisyydessä: SAADA on verbi, joten sen läheisyydessä esiintyy sen valenssiin kuuluvia jäseniä. Morfologista ja semanttista primingia rajoittaa vielä se, että niitä tutkitaan verbin tietystä merkityksestä. Tämä kyseinen merkitys on transitiivinen, mikä on omanlaisensa rajoite. Nämä seikat eivät kuitenkaan rajoita fraseologisia yksiköitä itsessään, vaan vain niiden esiintymisen tässä tutkimuksessa.

6) semanttisen yhtenäisyyden ja epäkompositionalisuuden/ennustamattomuuden rooli määritelmässä

Viimeinen parametri käsittelee fraseologisen yksikön semanttista yhtenäisyyttä (*semantic unity*) ja semanttista epäkompositionalisuutta/ennustamattomuutta (*semantic non-compositionality / semantic non-predictability*). Ilmiössä on kyse siitä, että elementit yhdessä saavat eri tulkinnan kuin elementit, jotka sen muodostavat. (Gries 2008: 4–6.) Tällaista ilmiötä havainnollistaa ilmaus *ei ole kaikki muumit laaksossa*, joka yhtenäisenä semanttisena kokonaisuutena kuvaa jonkinlaista mielenvikaisuutta, eikä suinkaan muumien sijaintia (ks. Kortelainen 2011: 5). Perinteisesti kuudes teesi on ollut tärkeää fraseologiselle tutkimukselle, sillä kiinnostus on kohdistunut edellä kuvatun tapaisiin leksikaalisiin yksikköihin, joiden merkitys on erilainen kuin sen yksittäisten elementtien summa (Granger & Paquot 2008: 31). Epäkompositionalisuudessa on kyse idiomeista, jotka ovat kuitenkin vain osa fraseologiaa.

Tässä tutkimuksessa noudatetaan kuitenkin uudempaa linjaa muun muassa Grangerin (1998) sekä Jantusen ja Brunnin (2012) tapaan ja käytetään laajentunutta fraseologisen yksikön käsitettä. Semanttinen epäkompositionalisuus ei ole siis enää kriteerinä, vaan se mielletään vain joillakin fraseologisilla yksiköillä olevaksi ominaisuudeksi (Granger & Paquot 2008: 31). Epäkompositionalisuus vaikuttaa tähän tutkimukseen kuitenkin siten, että mikäli osiensa summaa laajempia ilmauksia tulee vastaan, ne analysoidaan semanttisen primingin kokonaisuuden muodostaman merkityksen (aiemmin mainitut esimerkit *Vladimir Putin ja tänä päivänä*), ei yksittäisten elementtien mukaan.

2.2.3 Leksikaalinen priming

Aiemmin on ajateltu, että kielioppi on ollut ensin, ja sanat ovat ikään kuin pudotettu kieliopillisiin laatikoihin – tai että semantiikka syntyi ensin, ja leksikko aktualisoi semantiikan. (Hoey 2005: 1.) Hoey (mp.) kääntää päälaelleen perinteisen ajattelun sanaston ja kieliopin suhteesta: uuden ajattelutavan mukaan leksikko on rakenteeltaan kompleksinen ja systemaattinen, ja kielioppi on tämän leksikaalisen rakenteen tulos.

Hoeyn (2005: 7) mukaan priming on psykologinen ilmiö, jossa tietyn sanan kuuleminen herättää jonkin kohdesanan. Hoey (mts. 8) antaa tästä havainnollistavan esimerkin:

For example, a listener, previously given the word body, will recognize the word heart more quickly than if they had previously given an unrelated word such as trick; in this sense, body primes the listener for heart.

Primingia voi verrata ihmisen ja sosiaalisten rakenteiden väliseen suhteeseen, jossa jokainen tuottaa rakenteita uudelleen ja rakenne muokkaa yksilöä (Giddens 1979: 24; Hoey 2005: 8). Vuorovaikutustilanteessa puhujat ovat vapaita tiettyjen rajoitteiden puitteissa (Giddens 1979: 26): priming saa puhujan tarkoittamattaan toistamaan joitain kielen aspekteja, ja tämä uudelleen toistettu aspekti ”virittää”⁷ kuulijan (Hoey 2005: 9). Jokainen sana on siis virittynyt tietyssä diskurssissa käytettäväksi kumulatiivisten vaikutusten vuoksi, joita yksilö on maailmassa kohdannut (mts. 13).

Leksemeissä on koodattuna erilaisia ko- ja kontekstuaalisia ”käyttöohjeita” (Jantunen 2009b: 360) ja jokaisella sanalla on oma kielioppinsa sekä omat lokaaliset semanttis-pragmaattiset suhteensa (Jantunen 2009b: 360 mukaan Tognini-Bonelli 2008). Sinclairin (1991: 108) mukaan monilla sanoilla ei edes ole itsenäistä merkitystä, vaan ne ovat laajemman tekstin ja monisanaisten yksiköisen komponentteja, joista teksti muodostuu. Lisäksi leksikko, syntaksi ja semantiikka on mahdotonta (ja mieletöntä) erottaa toisistaan (mp; ks. Ullmann 1967: 237–238), koska merkityksen todellinen luonne ja toteutuminen kielenkäytössä on paljon yksityiskohtaisempaa kuin mitä perinteisesti kieliopissa on esitetty.

Hoey (ks. 2005: 13) esittää kymmenen väitettä sanojen primingista. Hoeyn (mp.) esittää leksikaalisen primingin kuvaukseksi seuraavaa:

⁷ 'Virittäytymisestä' merkityksessä 'prime' ovat aiemmin puhuneet suomeksi mm. Jantunen ja Brunni (2012) sekä Lindberg (2017).

1. sanat ovat virittäytyneet (*primed*) esiintymään toisten sanojen kanssa (kollokaatio, *collocation*),
2. sanat ovat virittäytyneet esiintymään tiettyjen merkitysryhmien kanssa (semanttinen assosiaatio, *semantic association*),
3. sanat ovat virittäytyneet esiintymään tiettyjen pragmaattisten funktioiden yhteydessä (pragmaattinen assosiaatio, *pragmatic association*),
4. sanat ovat virittäytyneet esiintymään tietyissä kieliopillisissa sijoissa (*grammatical positions*) tai funktioissa (*grammatical functions*) tai välttämään tiettyjä kieliopillisia sijoja tai funktioita (kolligaatio, *colligation*),
5. kohyponyymit (*cohyponyms*) ja synonyymit (*synonyms*) eroavat toisistaan kollokaatioiden, semanttisten assosiaatioiden ja kolligaatioiden osalta,
6. kun sana on polyseeminen, tietyn merkityksen kollokaatiot, semanttiset assosiaatiot ja kolligaatiot eroavat sanan muiden merkitysten kollokaatioista, semanttisista assosiaatioista ja kolligaatioista,
7. jokainen sana on virittäytynyt käytettäväksi yhdessä tai useammassa kieliopillisessa roolissa (kieliopillinen kategoria, *grammatical category*),
8. sanat ovat virittäytyneet esiintymään tietyn tyyppisissä koheesiosuhteissa (*cohesive relations*) tai välttämään tiettyjä koheesiosuhteita tietyissä diskursseissa (*a discourse*) (tekstuaalinen kollokaatio, *textual collocation*),
9. sanat ovat virittäytyneet esiintymään tietyissä semanttisissa suhteissa (*semantic relations*) tai välttämään tiettyjä semanttisia suhteita tietyssä diskurssissa (*the discourse*) (tekstuaalinen semanttinen assosiaatio, *textual semantic association*) ja
10. sanat ovat virittäytyneet esiintymään tietyissä diskurssin positioissa (tekstuaalinen kolligaatio, *textual colligation*). (Hoey 2005: 13.)

Kohdat kaksi (semanttinen assosiaatio), kuusi (polyseemisen sanan eri merkitysten erot) ja seitsemän (kieliopillinen kategoria) liittyvät suoraan tähän tutkimukseen, ja sen vuoksi vain niitä käsitellään tarkemmin. Kohdan kaksi ydin on, että jokainen sana on virittynyt esiintymään tiettyjen semanttisten joukkojen kanssa; tämä on sen semanttista assosiaatiota (primingia). Hoeyn (2005: 16–17) mukaan semanttinen assosiaatio kuvaa kielellistä luovuutta, mihin kollokaatio ei pysty. Esimerkkinä Hoey käyttää ajallista merkityskenttää: Englannin kielen tuntia merkitsevä sana *hour* tuntuu kollokoivan sanan *ride* (suom. ’ajaa, matkustaa’) kanssa, mutta se ei tunnu kollokoivan tarkkojen ajanmääreiden, kuten esimerkiksi *half an, one, two, three, four* ja *twenty four* kanssa, vaikka tuntuisikin loogiselta, että *hour* olisi natiivilla englannin kielen puhujalla virittynyt juuri ajanmääreisiin. Sanan *hour* semanttinen assosiaatio olisi siis aika, ja tähän ei pelkän kollokaatioanalyysin avulla pääsisi kiinni – koska *hour* myötäesiintyy useiden eri ajanilmausten kanssa, ei yksikään välttämättä esiinny tarpeeksi taajasti, jotta se näkyisi kollokaateissa. (Mp.)

Kohdan kuusi ydin on, että kun sana on polyseeminen, sen yhden merkityksen kollokaatiot, semanttinen priming ja kolligaatiot eroavat sen muista merkityksistä. Sinclair (1987) huomautti ensimmäisenä, että sanan jokaisella merkityksellä on omat kollokaatiot ja tavat tulla käytetyksi. Hoeyn (2005: 81) mukaan kielenkäyttäjät ovat virittyneitä tunnistamaan sanan rinnakkaisia ”kaavoja” ja erottamaan yhden merkityksen toisesta. Hoey (mts. 80) käyttää esimerkkinä elokuvan *Airplane vitsiä*, jossa pilotilla sanotaan olevan ’drinking problem’ (*alkoholiongelma*). Seuraavaksi näytetään pilotti kaatamassa alkoholitonta juomaa päälleen: ongelma

onkin se, ettei juoma osu hänen suuhunsa. Vitsi piilee siis sanajärjestyksessä. Jos pilotilla olisi kerrottu olevan 'problem drinking', joka on psykologinen häiriö, vitsi olisi pikemminkin surullinen kuin huvittava. Vaikka sanat *drinking* ja *problem* kollokoivat keskenään kummassakin tapauksessa, on olemassa vain yksi kieliopillinen yhdistelmä, että jollain on ongelma saada juomista kurkusta alas. Ei ole siis sattumaa, että kielenkäyttäjällä on usein tarve ja monia keinoja puhua alkoholismista, kun taas harvemmin tulee puhuttua psykologisesta juomishäiriöstä, ja siksi sille on vain yksi tapa. Yleisempi merkitys 'alkoholiongelma' ajaa harvemman merkityksen "kieliopilliseen nurkkaan".

Hoeyn (2005: 81) esimerkki osoittaa, että sanojen eri merkityksillä on kieliopillinen rakenne (ja myös semanttinen ympäristö). Edellä esitetyn esimerkin pohjalta Hoey esittää, että jos polyseemisen sanan taaja merkitys on virittäytynyt suosimaan tiettyjä kollokaatioita, semanttisia assosiaatioita ja/tai kollokaatioita, saman sanan harvempaan esiintyvä merkitys on virittäytynyt välttämään kyseisissä kollokaatioita, semanttisia assosiaatioita ja kolligaatioita. Taajempi merkitys saattaa kuitenkin toisinaan käyttää hyväkseen harvempaan esiintyvän muodon kontekstuaalista virittäytymistä. Sen sijaan jos sanan kaksi merkitystä ovat yhtä taajoja, molemmat ovat virittäytyneet välttämään toistensa kollokaatioita, semanttisia assosiaatioita ja/tai kolligaatioita. Jos nämä säännöt eivät päde, tuloksena syntyy huumoria, monitulkintaisuutta tai näistä kahdesta merkityksestä syntyvä uusi merkitys. (Mp.)

Polysemia ei muutenkaan ole ongelmatonta, sillä on hyvin vaikeaa määritellä, milloin merkitykset ovat erillisiä, ja montako merkitystä sanalla tarkalleen on (Palmer 1981: 100). Etenkin suomen kielen osalta polyseemisten sanojen eri merkitysten fraseologinen tutkimus on vähäistä, joten tämä tutkimus jo pelkästään siksi paikkaa tutkimusaukkoa.

Kohdan seitsemän ydin on, että jokainen sana on virittynyt käytettäväksi yhdessä tai useammassa kieliopillisessa roolissa; nämä ovat sen kieliopilliset kategoriat. Hoeyn (2005: 159) mukaan myös tavut (tai suomen kielen kontekstissa paremminkin morfeemit), kuten *ing*, *to* ja *ful* ovat virittäytyneet ja niilläkin on omat kollokaationsa, kolligaationsa ja semanttiset assosiaationsa. Agglutinoivassa kielessä, kuten suomen kielessä, tärkeä osa kielen kuvausta on olennaisten morfeemien primingin luonne. Vaikka Hoey ei asiaa samoin sanoin ilmaisekaan, ilmiö liittyy morfologiseen primingiin (ks. Jantunen & Brunni 2012). Hoeyn seitsemäs väite liittyykin vahvasti siihen, miksi tässä tutkimuksessa tarkastellaan SAADA-verbin morfologisen ja semanttisen primingin välistä suhdetta: morfeemit kantavat omaa merkitystään ja esiintyvät omissa merkitysympäristöissään. Seuraavissa alaluvuissa käsitellään Hoeyn (2005: 13) seitsemänteen väitteeseen liittyvää morfologista ja toiseen väitteeseen liittyvää semanttista primingia.

2.2.4 Morfologinen priming

Morfologia, eli muoto-oppi, tutkii sanojen sisäistä rakennetta niiden morfeemien kannalta; morfeemi on kielen pienin merkityksellinen yksikkö. Morfeemien merkitysten ja funktioiden tutkiminen yhdistää morfologian myös semantiikkaan ja syntaksiin. (Karlsson 1994: 75.) Edellä mainitut fraseologisen yksikön piirteet ovat olleet syntagmaattisia, mutta morfologinen priming on sekä syntagmaattinen (kontekstuaalinen) että paradigmaattinen. Kontekstuaaliseen ulottuvuuteen kuuluvat lemman eri sananmuotojen suhde kontekstiin sekä lemman suhde kontekstissa olevien kollokaattien eri sananmuotoihin. Paradigmaattinen ulottuvuus sisältää hakusanan sisäiset morfologiset valinnat. (Jantunen 2004: 28–29.) Tässä tutkimuksessa molemmat ulottuvuudet ovat läsnä. Ensin paradigmaattisesti selvitetään SAADA-verbin sisäiset morfologiset valinnat, minkä jälkeen tutkitaan eri sananmuotojen ja semanttisten luokkien välistä korrelaatiota, jolloin kyse on kontekstuaalisesta ulottuvuudesta, tarkemmin lemman sananmuotojen suhteesta kontekstiin.

Stubbs (2001) on tarkastellut kontekstuaalista morfologista primingia, lemman eri sananmuotojen kollokointia, SEEK-verbin (’etsiä, hankkia, tavoitella’) avulla. Joillakin muodoilla voi olla yhteisiä kollokaatteja (esim. *seek, seeking, sought*), mutta kaikilla muodoilla ei ole yhteisiä frekventtejä kollokaatteja (esim. *seeking, seeks*) (mts. 28). Tämä kertoo siitä, että eri sananmuodoilla on erilaisia konteksteja ja käyttöyhteyksiä. Jantunen (2001: 185) antaa paradigmaattisesta morfologisesta primingista esimerkiksi lähisyonymiset adjektiivit TÄRKEÄ ja KESKEINEN joista TÄRKEÄ on usein partitiivissa (*tärkeää*) ja KESKEINEN nominatiivissa (*keskeinen*). Myöhemmin Jantunen (2009b: 371–373) tarkasteli synonyymiparia HIEVAHTAA ja LIIKAHTAA. Selvisi, että LIIKAHTAA esiintyy usein finiittisenä ja imperfektissä, kun taas HIEVAHTAA esiintyy MATTA-infinitiivissä. Tämä osoitti, että HIEVAHTAA tarkoittaa usein liikkumattomuutta, kun taas LIIKAHTAA ilmaisee sekä liikettä että liikkumattomuutta. Nämä esimerkit havainnollistavat, kuinka kontekstuaalinen morfologinen preferenssi erottelee saman lemman eri sananmuotoja ja kuinka paradigmaattinen morfologinen priming erottelee (lähi)synonyymejä toisistaan. Morfologista primingia onkin syytä tutkia lisää, jotta sanojen (ja sananmuotojen) välisistä suhteista saataisiin tarkempi kuva. Tässä tutkimuksessa paradigmaattinen morfologinen preferenssi erottelee eri kielimuotojen SAADA-verbin käyttöä ja kontekstuaalinen semanttinen preferenssi erottelee eri sananmuotoja kummankin kielimuodon sisältä.

Sinclairin (1991: 103) mukaan kieliopissa ei ole huomioitu sanastoa ja merkitystä – esimerkiksi sitä, että tietty kieliopillinen malli voi olla ominainen vain yhdelle sanan useasta merkityksestä. Saman sanan eri merkityksillä voi olla siis tendenssi esiintyä aivan eri muodoissa

kuin muut merkitykset. Esimerkiksi käy omakohtainen tutkimusavustajakokemus (tutkimuksessa Jantunen 2018) jossa lemmän VAMMAINEN sananmuoto *vammaisia* esiintyi halventavana adjektiivina, kun taas substantiivisella *vammaiset*-muodolla puhuttiin ihmisryhmästä. Sinclair (1991: 8) ja Stubbs (2001: 17) ovatkin yhtä mieltä siitä, että fraseologisia yksiköitä tulisi tarkastella lemmatasolla ainoastaan, mikäli lemmän sananmuotojen syntagmaattiset valinnat olisivat yhteneväiset⁸, koska yksittäisillä sananmuodoilla voi olla erilaisia käyttötapoja (mm. Stubbs 2001: 49).

Fraseologisen yksikön ytimellä ja sen kontekstuaalisilla elementeillä on taipumusta suosia tiettyä taivutusmuotoa (Jantunen & Brunni 2012: 73). Kuitenkin, koska fraseologisen tutkimuksen perinteet ovat englannin ja muiden indoeurooppalaisten kielten tutkimuksessa, morfologiset piirteet ovat jääneet vähäiselle huomiolle myös suomalaisessa fraseologisessa tutkimuksessa (etenkin natiivi- ja oppijankielen korpustutkimuksessa) (mts. 72–73), vaikka morfofonologisesti⁹ rikkaassa suomen kielessä morfologinen priming kuuluu olennaisesti kielen fraseologiseen luonteeseen (mts. 71, 75). Tästä syystä Jantunen ja Brunni (mts. 78) esittävät morfologista primingia osaksi Hoeyn (2005) leksikaalisen primingin teoriaa (ks. luku 2.2.3)¹⁰ seuraavasti (ks. myös Jantunen 2017: 358, 267–269):

Jokaisella sanalla on oma morfologinen priminginsa, eli se esiintyy tyypillisesti ja frekventisti yhdessä tai useammassa morfofonologisessa muodossa, ja myös sen kontekstissa esiintyvistä ilmauksista on löydettävissä omat odotuksenmukaiset taivutusmuotonsa.

Jantunen ja Brunnin (2012: 77) esimerkkiä noudattaen tässäkin tutkimuksessa käytetään käsitettä morfologinen priming käsitteen morfologinen preferenssi¹¹ sijaan.

Usein morfologista primingia on tutkittu synonyymipareilla (synonyymien eroavaisuuksien tarkastelussa), mutta tuon tutkimuksellani näkökulman saman verbin vertailuun natiivi- ja oppijankielessä yhden verbin merkityksen sisällä. Vertailu natiivi- ja oppijankielen välillä vastaa Jantunen ja Brunnin (2012) esiin nostamaan ongelmaan morfologisen primingin ja fraseologian syy-seuraussuhteesta: vertaileva asetelma paljastaa, missä kohti oppijat tekevät morfologisia epätyypillisyyksiä ja toistuvatko epätyypillisyydet myös semanttisesti. Morfologisen primingin ilmiötä synonyymiparein on tutkinut myös Arppe (2002), joka huomauttaa, että

⁸ Näin käy kuitenkin vain harvoin (Jantunen & Brunni 2012: 78).

⁹ Morfofonologisella vaihtelulla tarkoitetaan sitä, että morfeemiin kuuluvalla äännesegmentillä on erilaisissa ympäristöissä erilaiset foneemiedustumat (VISK § 40).

¹⁰ Hoey (2005: 159) on viitannut morfologisen primingin ilmiöön, mutta jättänyt sen ulos varsinaisesta leksikaalisen primingin teoriastaan.

¹¹ Morfologisesta primingista on aiemmin käytetty termiä morfologinen preferenssi (mm. Jantunen 2004; 2009b).

täydellinen synonymia on harvinaista ja nimenomaan kontekstin rajoitukset erottavat synonyymit toisistaan. Esimerkiksi Arpen (mp.) tutkimuksen mukaan MIETTIIÄ esiintyy taajemmin yksikön ensimmäisen persoonan muodossa kuin lähisynonyymi POHTIA.

Laajemmin fraseologian viitekehyksessä morfologisen primingin tarkastelu on tärkeää, koska morfologinen priming ja oppijankielen fraseologisuus ovat syy-seuraussuhteessa: kielenoppijoiden epätyypillinen fraseologisuus selittää osan natiivikielestä eroavasta morfologisesta primingista. (Jantunen & Brunni 2012: 71.) Morfologisen primingin tarkastelu on tärkeää myös siksi, että sananmuotojen tarkastelu itsenäisinä yksiköinä tarjoaa paljon lisätietoa sanojen ja sanastojen käyttäytymisestä, ja tietoa voi soveltaa esimerkiksi kieliopin ja sanastojen kuvauksissa sekä kielenopetuksessa. Lisäksi morfologinen priming on yhteydessä polyseemisen sanan eri merkityksiin (Jantunen & Brunni 2012: 71), mikä puolestaan motivoi tämän tutkimuksen polyseemistä otetta.

2.2.5 Semanttinen priming

Semanttinen priming, tai tutummin semanttinen preferenssi (terminologisiin valintoihin palataan tämän alaluvun lopussa) on keskeinen käsite fraseologisessa korpustutkimuksessa (Bednarek 2008: 119). Se tarkoittaa lekseemin kollokointia tietyn semanttisen ryhmän kanssa (mts. 119, 121). Lisäksi Jantunen (2009b: 359) määrittelee semanttisen preferenssin leksikaalis-semanttiseksi myötäesiintymäksi, joka on lekseemin ja kotekstissa olevan merkityspiirteen syntagmaattinen suhde (ks. kuvio 2). Toisin sanoen semanttinen priming tarkoittaa suhdetta sanan sekä sellaisen sanajoukon välillä, jotka liittyvät merkityksensä puolesta toisiinsa (Stubbs 2001: 65).

Sinclair (1998: 15–16) on ottanut termin ensimmäisenä käyttöön, ja hänen mukaansa (1991: 112) monella sanalla on taipumusta esiintyä tietyssä semanttisessa ympäristössä. Esimerkiksi englannin kielen verbi *happen* (suom. 'tapahtua') assosioituu epämiellyttävien asioiden kanssa, kuten kolareiden (mp.). Tätä Hoey (2005: 22) havainnollistaa sanonnalla *shit happens*, mikä hänen mukaansa tiivistää *happen*-verbin semanttisen luonteen. Semanttista preferenssiä tutkitaan merkityskeskeisesti esimerkiksi sanaluokkarajojen yli (Sinclair 1998: 16). Esimerkiksi Jantunen (2009b: 366) tutkimuksessa lekseemin AIVAN semanttinen preferenssi 'selvyys' näkyy sanoista SELVÄ, KIRKASTUA ja *selvästi*. Semanttinen preferenssi ei välttämättä ole myöskään positiosidonnainen, esimerkiksi lekseemin AIVAN semanttinen preferenssi 'paikka' ilmenee postpositionarakenteesta *oven vieressä* ja sanoista *lähelle*, *sängyssä* ja *rannoilla*. Hoey (2005: 37) huomauttaa, että semanttinen preferenssi ei kuitenkaan ole vapaa

kieliopillisista rajoitteista, mikä korostaa kontekstuaalisen valinnan läsnäoloa tässä tutkimuksessa. Huomautus perustele myös valintaa tutkia semanttisen preferenssin suhdetta morfologiseen primingiin.

Sinclairin jälkeen muutkin ovat esittäneet oman määritelmänsä semanttiselle preferenssille. Esimerkiksi Stubbsin (2001: 65) mukaan semanttinen preferenssi on suhde, joka vallitsee sanan (lemman tai sananmuodon) sekä semanttisesti siihen liittyvän sanajoukon välillä. Stubbs (mp.) antaa esimerkiksi sanamuodon *large* (suom. ’suuri’), joka esiintyy usein määrää ja kokoa ilmaisevien sanojen läheisyydessä. Pirkolan (2016) mukaan kyse on siitä, mihin merkitysryhmiin kuuluvia sanoja yleensä esiintyy tutkittavan sanan ympäristössä, eli millaisia semanttisia piirteitä sana saa kontekstissaan. Huomionarvoista on, että kyse on toistuvasta merkityspiirteestä (Sinclair 1998: 16; Stubbs 2001: 449). Hoey (2005: 17–18) käyttää termiä semanttinen assosiaatio, Louw (1993: 157) semanttista prosodiaa (mikä taas eri määritelmien mukaan voi tarkoittaa semanttista prosodiaa tai preferenssiä).

Bednarekin (2008: 122) mukaan lekseemin semanttisen preferenssin tunnistaminen on osin ongelmallista, sillä semanttisen preferenssin tai prosodian tunnistamiseksi tutkijan täytyy määrittää tarkasteltavan lekseemin semanttinen ryhmä, mitä on vaikeaa toteuttaa objektiivisesti. Joidenkin lekseemien semanttinen preferenssi on helposti määriteltävissä, mutta aina näin ei ole. Kangasniemi (1997: 16) vertaakin lekseemin merkityksen määrittelyn prosessia ”yritykseen määrittellä jokin äänne ääntämättä sitä”. Lisäksi lekseemeillä voi olla useampia semanttisia preferenssejä ja merkitysryhmiin mahtumattomia kollokaatteja (Jantunen 2004: 25), mikä vaikeuttaa semanttisten luokkien tunnistamista.

Partington (2004) pyrkii artikkelissaan selventämään semanttisen preferenssin ja prosodian välistä eroa. Semanttisen prosodian käsitettä ovat hänen mukaansa käsitelleet perusteellisesti Firthin jälkeen muun muassa Sinclair (1987, 1996, 1998), Louw (1993) ja Stubbs (2001). Tästä esimerkkinä on Sinclairin (1987) kuuluisa (ja jo aiemmin viitattu) esimerkki siitä, että englannin kielen sana *happen* esiintyy negatiivisessa ympäristössä. Semanttisen prosodian voisikin mieltää abstraktina (positiivisena tai negatiivisena) merkitysourana (Jantunen 2009b: 359), kun taas semanttinen preferenssi viittaa tarkempaan ja rajatumpaan merkitykseen.

Semanttinen preferenssi on todennäköisesti kieliyhteisön puhujien jakamaa tietoutta (Partington 2004: 152; Bednarek 2008: 122). Tästä näkökulmasta semanttista preferenssiä on viime aikoina alettu pitää sanan vireenä (*priming*), mikä tarkoittaa sitä, että kun sana on opittu käyttöyhteyksissä sekä puhuttuna että kirjoitettuna, se on ladattu näiden käyttöyhteyksien kumulointuneilla vaikutuksilla niin, että on osa sanan ymmärtämistä tietää, kuinka se myötäesiintyy muiden sanojen kanssa (ks. Hoey 2005, Bednarek 2008: 123). Tämän vuoksi tässä tutkimuksessa

käytetään ilmiöstä termiä semanttinen priming semanttisen preferenssin asemesta. Taustalla on sama ajatus kuin morfologisen primingin termin valinnan osalta: pyrkimys liittää ilmiö tiukemmin osaksi leksikaalisen primingin teoriaa. Hoey (2005) on sisällyttänyt semanttisen assosiaation teoriaansa, mutta semanttinen assosiaatio voitaisiin nähdä paremminkin yläkäsitteenä semanttiselle prosodialle, primingille ja muille mahdollisille semanttisille yksiköille laajan merkityksensä vuoksi. Esimerkiksi KS:n (s. v. *assosiaatio*) kyse on (mielle)yhtymästä, mikä liittyynee kaikkeen semantiikkaan.

Tässä tutkimuksessa semanttinen priming määritellään semanttisen kentän kautta. Semanttinen kenttä tarkoittaa sitä, että saman aihepiirin sanat muodostavat oman kentän tai luokan: esimerkiksi sukulaissanoista muodostuu semanttinen kenttä, koska kaikki sukulaista tarkoittavat sanat liittyvät jollain tasolla toisiinsa, esimerkiksi *äiti* ja *poika* liittyvät toisiinsa ‘perheen’ semanttisen kentän kautta (Kuiiri 2012: 38; esimerkki oma). Oleellista on, että saman kentän sanat assosioituvat toisiinsa, eli kenttä ei voi olla täysin mielivaltaisen. Muita esimerkkejä semanttisista kentistä ovat esimerkiksi liikkumista kuvaavat verbit ja jääkiekon sanasto. (Kuiiri 2012: 38.) Semanttinen priming nähdään myös tietynlaisena kenttänä, koska sanoja luokitellaan niiden merkitysten mukaan saman semanttisen kentän mukaisiin luokkiin. Semanttinen priming eroaa semanttisesta kentästä kuitenkin siten, että semanttinen priming on sidottu johonkin lekseemiin tai muuhun kokonaisuuteen (tässä tutkimuksessa SAADA-verbiin), kun taas semanttinen kenttä ei ole varsinaisesti sidoksissa muihin elementteihin. Voisi siis ajatella, että semanttinen priming on tietyn lekseemin semanttinen kenttä.

3 OPPIJANKIELI JA FRASEOLOGIA

3.1 Oppijankieli

3.1.1 Oppijankielen määritelmä ja oppimiskontekstit

Oppijankieli on kirjoitettua tai puhuttua kieltä, jota tuottaa kielenoppija (Ellis R. & Barkhuizen 2005: 4). Oppijankielen voi määritellä tarkoittamaan myös kielenoppijoiden yksilöllisiä kieli-järjestelmiä (joissa toki on myös yhteneviä universaaleja piirteitä, ks. luku 3.1.2) (Martin & Alanen 2011: 33; Nissilä 2011: 39). Nissilän (mp.) mukaan oppijoiden tuottama kieli on kielenoppimisprosessissa kehittyvä kielimuoto, ja myös Martin ja Alanen (2011: 33) kuvaavat kielenoppimista luonteeltaan prosessinomaiseksi. Oppijankieltä on kutsuttu myös välikieleksi (*interlanguage*) tai välivaihekieleksi ja oppijansuomea on kutsuttu myös ulkomaalaisuomeksi ja vierassuomeksi. Nissilän (2011: 37) tavoin tässä tutkimuksessa oppijankieltä pidetään omana kielimuotonaan enkä pelkästään kielenoppimisen välivaiheena, joten siitä käytetään termiä oppijankieli merkityksessä 'oppijan tuottama kieli' ja termiä oppijansuomi täsmällisemmässä merkityksessä 'oppijan tuottama suomi'. Termivalinta on perusteltu myös siitä näkökulmasta, että oppijankieli (*learner language*) on toisen ja vieraan kielen (L2) tutkimuksessa vakiintunut termi. Oppijankieli syntyy Nissilän (mts. 39) mukaan lähtökielen (kielenoppijan äidinkielen¹²), opittavan kielen (kohdekielen) ja oppijan kulloisenkin oppijankielen vaiheen ja oppijan osamien muiden kielten vaikutuksesta. Luonteeltaan oppijankieli on dynaaminen, alati muuttuva. Oppijankieli ei myöskään ole yksi homogeeninen kielimuoto, vaan pikemminkin joukko samankaltaisia kielisysteemejä. (Latomaa 1993: 20–21; Nissilä 2011: 39–40.)

Jotta voidaan määritellä oppijankieli (tai natiivikieli), täytyy myös pohtia kielen ole-musta. Tässä tutkimuksessa nähdään, että kieli on todella olemassa (eikä vain saussurelainen järjestelmä): kieli nähdään konkreettisenä vuorovaikutustoimintana¹³ abstraktin järjestelmän sijaan (ks. Dufva, Aro, Suni & Salo 2011: 31). Myös Mustonen (2015: 56) korostaa, että kieli kuuluu äidinkielisten lisäksi myös oppijoille eikä kielenoppija vastaanota valmista systeemiä, vaan on osana muokkaamassa sitä. Kielen oppimisen näkökulmasta tämä tarkoittaa sitä, että

¹² Tutkimuksessa tiedostetaan äidinkielen käsitteen ongelmallisuus, ja sitä käytetään vain neutraalisti tutkimustarpeisiin kuvaamaan L1-kieltä tai -kieliä, jo(i)ta henkilö osaa parhaiten ja on yleensä oppinut ensimmäisenä.

¹³ Vaikka kielenopetuksen traditio painottuu kieliopin opettamiseen, muun muassa Corder painotti viestinnällisyyden merkitystä kielenopetuksessa jo 70-luvulla (Corder 1976: 26).

sosiaalinen ja kognitiivinen toiminta nivoutuu yhteen (mts. 30). Dufva ynnä muut (mp.) kiteyttävät asian seuraavasti:

[O]ppimista on tärkeää tarkastella toimintana erilaisissa kieliympäristöissä, ja oppiminen kognitiivisella tasolla tapahtuu nimenomaan osallistumalla sosiaaliseen toimintaan.

Tässä tutkimuksessa oppijankielen nähdään siis muodostuvan sellaisten henkilöiden konkreettisista suomenkielisistä tuotoksista, joiden äidinkieli ei ole suomi. Kirjallisetkin tuotokset nähdään omana modaliteettinaan osana vuorovaikutustilanteita eikä kuten perinteisesti on ajateltu, edustamassa koko kieltä (ks. Dufva ym. 2011: 24, 30).

Oleellista on myös pohtia, kuka on kielenoppija ja milloin kielenoppiminen lakkaa – vai lakkaako ollenkaan? Tähän kysymykseen vastataksaan täytyy ensin miettiä oppimisprosessin vaiheita alusta alkaen. Kielenoppijat voi jakaa perustason kielenkäyttäjiin (A1–A2), itsenäisiin kielenkäyttäjiin (B1–B2) ja taitaviin kielenkäyttäjiin (C1–C2) Eurooppalaisen viitekehyksen (EVK) mukaan (EVK 2003: 47). Tässä tutkimuksessa oppijansuomen aineisto koostuu taitotasoina B2–C2. EVK:n luokittelun mukaan oppijat voi jakaa taitotasoihin muun muassa kielen oppimiseen käytetyn ajan mukaan (ks. Nissilä 2011).

Yksi tapa hahmottaa kielen oppimista taitojen mukaisesti on Corderin (1973: 275–276) jako kielenoppimisen neljään vaiheeseen:

- 1) Harjoitusvaihe, jossa virheet ovat umpimähkäisiä. Oppija voi näyttää osaavan säännön, mutta ei osaa soveltaa sitä, minkä vuoksi välillä oppija onnistuu ja välillä taas tekee virheitä.
- 2) Presystemaattinen vaihe, eli järjestelmällisyyttä edeltävä vaihe. Tällä tasolla virheet ovat sattumanvaraisia, eli oppija tuottaa edelleen satunnaisesti oikeita ja virheellisiä muotoja. Vaihetta voisi luonnehtia niin, että oppijalla on vain epämääräinen käsitys (jos käsitystä lainkaan) siitä, että hänellä on erityinen järjestelmä opittavanaan kohdekielessä.
- 3) Järjestelmällisten virheiden vaihe, jolloin kielenoppija alkaa hahmottaa sattumanvaraisilta näyttävyyville ilmeille funktioita. Tässä vaiheessa oppija uskaltaa kokeilla erilaisia kielen järjestelmiä ja hypoteeseja.
- 4) Postsystemaattinen vaihe, jossa oppija voi vielä hetken soveltaa oppimaansa epäjohdonmukaisesti, mutta jossa oppija keksii oikean järjestelmän.

Kaikki kielen osa-alueet eivät kehity samaan aikaan, vaan ne voivat olla tasoilla eri aikoihin. On myös hyvä muistaa, että arvaamisella on paikkansa kielen oppimisessa: arvaukset voivat olla aluksi umpimähkäisiä ja myöhemmin perusteltuja ja järjestelmällisiä. (Mts. 275.) Toisaalta kehitys saattaa myös pysähtyä ennen kohdekielen hallinnan saavuttamista: tätä kutsutaan fossilisaatioksi. Tällöin jatkuvasta kohdekielen syötöksestä huolimatta tuotokset pysyvät virheellisinä. (Mts. 274.) Koska tämän tutkimuksen oppijansuomen aineisto on kerätty itsenäisen ja taitavan kielenkäyttäjän taitotasoilta B2–C2 (EVK 2003: 47), on oletettavaa, että oppijat sijoittuisivat suurimmaksi osin myös Corderin (1973) luokittelun yläpäähän.

Täytyy kuitenkin muistaa, että tällainen jako on teoreettinen, koska kielenoppija on samanaikaisesti myös kielenkäyttäjä. Natiivipuhujatkin ovat omalla tavallaan kielenoppijoita, koska kielen täydellinen hallinta on saavuttamaton abstraktio, minkä vuoksi kielenoppiminen ei lopu koskaan. (Mustonen 2015: 53.) Tässä tutkimuksessa on kuitenkin mielekästä erotella terminologisesti äidinkieliset kielenkäyttäjät (natiivit) suomi vieraana kielenä -kielenkäyttäjistä (kielenoppijoista) vertailuasetelman vuoksi.

Toisen kielen oppimisella voidaan tarkoittaa yläkäsitteenä kaikkea muuta kuin äidinkielen kielenoppimistilannetta (Ellis R. 1985: 5). Kansainvälisesti tästä tutkimusalasta käytetään lyhennettä SLA (*Second Language Acquisition*) (Nissilä 2011: 38). Toinen kieli (*second language*) voi tarkoittaa myös kieltä, jota opitaan omaksumalla kieliyhteisön parissa. Vierasta kieltä (*foreign language*) sen sijaan opitaan kohdekielisen kieliyhteisön ulkopuolella. (Kaivapalu 2005: 20; Nissilä 2011: 37–38.) Tämän tutkimuksen aineiston oppijankielen osuus koostuu juuri suomi vieraana kielenä -oppijoiden teksteistä, jotka on tuotettu ulkomaisissa yliopistoissa.

Latomaan ja Tuomelan (1993: 243) mukaan äärimmillään vieraan kielen oppiminen on “ulkomailla vain opetuksen kautta opittavaa suomea” ja toisen kielen oppiminen on Suomessa nimenomaan “kadulla omaksuttavaa suomea”. Nissilä (2011: 38) kuitenkin korostaa, että täydellistä jakoa vieraan ja toisen kielen oppimiselle ei voi tehdä, vaan kyse on jatkumosta. Toisaalta oppijankontekstin määrittely on ongelmallista, sillä kontekstit eivät ole yksiselitteisiä, vaan niissä tulisi ottaa huomioon useita muuttujia oppijan äidinkielestä käytettyihin oppimateriaaleihin (Mukherjee & Götz 2015: 423). Tämän lisäksi kielenoppija voi kielenoppimisprosessin aikana kuulua useampaan eri oppimiskontekstiin, eli konteksti voi myös muuttua (Mukherjee & Götz 2015: 440).

Ellis R:n (1985: 5) mukaan oppijankielen tutkimus perustuu ensimmäisen opittavan kielen (äidinkielen) tutkimukseen, ja keskeistä oppijankielen tutkimuksessa onkin, millä tavoin äidinkielen ja toisen ja vieraan kielen oppiminen eroavat toisistaan, ja mitä yhteistä niillä on. Grangerin (1998: 4) mukaan oppijankielen tutkimuksen tärkein tavoite on kartoittaa periaatteita, jotka ohjaavat toisen ja vieraan kielen oppimisprosessia. Vertailevalla tutkimuksella pyritään selvittämään kielenkäytön eroja äidinkielisten ja kielenoppijoiden välillä, jotta päästäisiin kiinni systemaattisesti toistuviin eroavaisuuksiin sekä pohtimaan eroja aiheuttavia tekijöitä: täytyy kuitenkin muistaa, että aineisto antaa pääsyn vain oppijankielen tuotoksiin, ei varsinaiseen prosessiin.

3.1.2 Oppijankieli universaalina kielivarianttina

Oppijankielestä on paljon tutkimuksia, mutta myös paljon sirpaleista tietoa. Jantunen (2008) on koonnut tietoa yhteen ja esittänyt hypoteeseja oppijankielen universaaleista. Oppijankielen universaalilla tarkoitetaan oppijankielen yleisiä piirteitä lähtö- ja kohdekielestä riippumatta (mts. 3).

Ensimmäinen hypoteesi on, että oppijankielessä kielenaineiksilla on epätyypillisiä taajuuksia (Jantunen 2008: 4). Oppijankielestä osataan usein sanoa, että jokin piirre yliedustuu suhteessa natiivikieleen, mutta harvemmin tutkimus keskittyy siihen, mitkä piirteet aliedustuvat (Ellis R. & Barkhuizen 2005: 70; Jantunen 2008: 4 mukaan Dagneaux, Dennes & Granger 1998: 163). Tämä johtunee virheanalyysin vahvasta traditiosta, jossa huomio on keskittynyt oppijoiden tekemiin virheisiin: tietyn rakenteen vähäiset virheet eivät kuitenkaan välttämättä johdu rakenteen hallinnasta, vaan pyrkimyksestä välttää sitä (Ellis & Barkhuizen 2005: 70).

Oppijankielessä yli- ja aliedustuminen voi koskea sekä sanastollisia että kieliopillisia kielenaineiksia (Jantunen 2008: 5). Esimerkiksi virolaisilla suomenoppijoilla yliedustuvat demonstratiivipronominin *se* ja *ne*, sillä niitä käytetään demonstratiivien *tämä*, *nämä*, *tuo* ja *nuo* asemesta (Heikkilä-Kopperoinen 2004: 99). Lisäksi suomenoppijoilla OLLA-verbi on yliedustunut, sillä sen avulla korvataan vaikeampia verbejä (Grönholm 1993: 144–145). Kansainvälisesti on havaittu, että lähtökielestä riippumatta englanninoppijat käyttävät puolet vähemmän *ing-* ja *ed-*muotoisia partisiippirakenteita natiiveihin englannin käyttäjiin verrattuna (Granger 1997: 3–5, 7, 9).

Poikkeavia taajuuksia voi selittää kielenaineiksen yleistäminen epätyypillisiin käyttöyhteyksiin tai se, että oppija alkaa käyttää vasta opittuja sanoja ja rakenteita juuri oppimisen jälkeen (Latomaa 1996: 99). Tällöin oppimisprosessi näkyy oppijan tuotoksessa ja sitä kautta tutkimusaineistossa (Jantunen 2008: 6). Poikkeavat taajuudet voivat selittyä osin myös ensikielen vaikutuksella, vaikka yli- ja alikäyttöä todennäköisesti selittääkin useampi tekijä (Ellis R. 1994: 305–306).

Tästä päästäänkin toiseen hypoteesiin, kieltenväliseen vaikutukseen (myös siirtovaikutus, interferenssi tai transfer). Yleisesti ajatellaan, että oppijan äidinkieli vaikuttaa opittavaan kieleen. (Jantunen 2008: 6–7; ks. 1993: 23–25; Ellis R. 1985: 19; 1994: 299–345.) Äidinkieli ei ole kuitenkaan ainoa vaikuttava kieli: ensimmäinen opittava vieras kieli saa vaikutteita oppijan äidinkielestä, kun taas toinen opittava vieras kieli saa vaikutteita enemmän ensimmäisestä vieraasta kielestä kuin oppijan omasta äidinkielestä (Dewaele 1998: 486, 488).

Kielten välisen vaikutuksen tutkimus on keskittynyt erityisesti indoeurooppalaisiin kieliin ja syntaksin, semantiikan ja fonologian ilmiöihin (Ellis R. 1985: 19; Kaivapalu 2005: 17, 27). Kielten välistä vaikutusta on todennettu myös suomalais-ugrilaisissa kielissä. Esimerkiksi Kaivapalu (2005) osoittaa, että kohdekielen (suomi) sanojen taivuttamisessa on apua, jos lähdekieli (viro) on kohdekielen sukukieli. Spoelman (2013) puolestaan todistaa, että vironkielillä oppijoilla on ei-sukukielitaustaisia oppijoita vähemmän partitiivivirheitä suomen kielessä.

Kolmas hypoteesi koskee oppijankielen yksinkertaisuutta. Onkin odotustenmukaista, että erityisesti kielenoppimisen alkuvaiheessa oppijankieli on monella tavalla yksinkertaisempaa kuin aikuisen natiivin kieli. On myös tärkeää huomata, ettei yksinkertaisuus ole tässä yhteydessä arvottava ilmaus, vaan se on luonnollisena osana oppijankieleen kuuluva piirre eri kehityksen vaiheissa. (Jantunen 2008: 8–9.) Yksinkertaistaminen on yksi kielenoppijan strategisista keinoista viestintätilanteessa, ja se näkyy pyrkimyksenä välttää monimutkaisia rakenteita ja selvittää jo hallituilla keinoilla virheiden välttämiseksi. Muita oppijankielen yksinkertaisuuteen johtavia tekijöitä ovat esimerkiksi yläkäsitteiden ja kiertoilmausten käyttö. (Jantunen 2008: 9 mukaan Faerch & Kasper 1984: 48–51.) Edellä mainittuja strategioita hyödynnetään erityisesti puhetilanteessa, mutta myös kirjoitetussa viestinnässä. Kirjoittaessa oppijalla on tosin enemmän aikaa miettiä keinoja ja käyttää apuvälineitä. Kovin tarkkaa tietoa, millaisina piirteinä yksinkertaisuus todellisuudessa konkretisoituu oppijankielessä, ei kuitenkaan ole. (Jantunen 2008: 9–10.)

Muutamia esimerkkejä yksinkertaistamisesta on kuitenkin havaittu: Oppijantuotoksessa taajimmin käytetyt adjektiivit muodostavat suuremman osan koko adjektiivikirjosta kuin natiivipuhujilla, mikä viittaa adjektiivien yksinkertaiseen käyttöön oppijankielessä (Nieminen 2001: 98–100). Lisäksi on viitteitä oppijankielen syntaktisesta yksinkertaisuudesta (ks. Hyland & Milton 1997: 385). Esimerkiksi Jantunen (2017: 261–262, 268) osoittaa, että oppijankielessä suositaan *kello*-ilmauksissa yksinkertaista rakennetta (*kello kaksi*) kompleksisemmän (*kello kahdelta*) sijaan. On myös osoitettu, että oppijat turvautuivat tuttuihin sanoihin, ”leksikaalisiin nallekarhuihin” (*lexical teddybears*), mikä johtaa tuttujen muotojen yliedustumiseen (Hasselgren 1994: 250, 256–257). Lisäksi tiedetään muun muassa, että suomenoppijat käyttävät vain osaa kaikista mahdollisista astemääreistä, mikä myös tukee ajatusta kielenoppijoiden yksinkertaisista tuotoksista (Jantunen 2015: 113). Oppijankielen yksinkertaisuudella ja natiivikielestä poikkeavilla kielenainesten taajuuksilla näyttäisi olevan siis jonkinlainen yhteys. On kuitenkin huomattava, että on myös ristiriitaisia tutkimustuloksia oppijankielen yksinkertaisuuden suhteen (Granger & Wynne 1998: 7). Ristiriitaisuus johtunee osin erilaisista tutkimusmetodeista, ja myös tutkimusaineisto on voinut aiheuttaa tulkinnallisia ongelmia (Jantunen 2008: 12).

Neljäs hypoteesi on, että oppijankieli on epäkonventionaalista. Hypoteesi sisältää kaksi erilaista näkökulmaa: ensimmäinen koskee oppijankielen yleiskielisyyttä ja puhekielisyyttä, toinen kontekstuaalista epäkonventionaalisuutta. On luonnollista, että yleis- ja puhekielen normistot sekoittuvat oppijankielessä, koska kielenoppijoilla ei ole varsinkaan kielenoppimisen alkuvaiheessa taitoa hallita erilaisten kielimuotojen monimutkaisia säännöstöjä. (Jantunen 2008: 12–13.)

On oletettavaa, että kielenoppijan puhekieli on yleiskielisempää kuin natiivipuhujalla (Jantunen 2008: 13). Maahanmuuttajien kielessä vain harvat yleiset puhekielen piirteet ovat tulleet osaksi kielenoppijan puhekieltä. Tällaisia ovat kuitenkin ensimmäisen ja toisen persoonan pikapuhemuodot *mä : mun, sä : sun*) ja possessiivisuffiksin puuttuminen (*mun kirja*) (Jantunen 2008: 13 mukaan Maijanen 1993: 69) sekä sananloppuisen *i:n* heittyminen (*yks*) (Jantunen 2008: 13 mukaan Mikkonen 2007: 88). Lisäksi suomenoppijat käyttävät adjektiivien yhteydessä mieluummin yleiskielisiä kuin puhekielisiä määritteitä (Nieminen 2001: 109). Suomenoppijoiden omaksumat puhekieliset piirteet näyttäisivät olevan yleisiä puhekielisyyksiä, eivätkä esimerkiksi murteellisia piirteitä (Jantunen 2008: 13–14 mukaan Mikkonen 2007: 84–87). Puhutun oppijankielen yleiskielisyyttä selittää osaltaan opettajien käyttämä yleiskielinen opetuspuhe ja oppikirjojen runsas yleiskielisyys (ks. Storhammar 1994: 101–103). Myös oppimiskonteksti vaikuttanee: toisen kielen oppimiskontekstissa voi olla helpompaa omaksua puhekielisyyksiä kuin kieliyhteisön ulkopuolella vieraan kielen oppimiskontekstissa.

Epäkonventionaalisuutta puhe- ja yleiskielen suhteen esiintyy myös kirjoitetussa oppijankielessä (Jantunen 2008: 14). Esimerkiksi englanninoppijoiden esseekirjoituksissa esiintyy puhekielisiä piirteitä, kuten puhekielinen *I think* -ilmaus sekä puhekieliset sanat *maybe*, *really* ja *absolutely* (Gilquin & Paquot 2008: 6–8). Oppijansuomesta on havaittu, että ylioppilasaineissa puuttuvat puhekielen mukaisesti omistusliitteet (*meidän ajatuksia*) sekä subjektin ja predikaatin kongruenssi (*me viestittäään*) (Jantunen 2008: 15 mukaan Manninen 1997: 73, 75; ks. myös Lähdemäki 1995).

Viimeisen hypoteesin toinen osa koskee kontekstuaalista epäkonventionaalisuutta eli tekstuaalista, kontekstuaalisen merkitysyksikön rakenteellista epätyypillisyyttä. Sen voi kuitenkin ajatella liittyvän myös laajemmin kontekstuaalisiin ominaisuuksiin, esimerkiksi tyyli- ja rakennepiirteiden hallintaan. (Jantunen 2008: 15.) Kontekstuaalisella merkitysyksiköllä tarkoitetaan kielellisten yksiköiden ja niiden käyttösääntöjen kokonaisuutta, johon sisältyy valintarajoitteita ja -preferenssejä (Jantunen 2004: 15, 19). Oppija joutuu kokeilemaan kielellisten yksiköiden sovittamista konteksteihin ilman tietoa kohdekielen tyypillisistä lekseemiä laajemmista ja monisyisistä ”elementeistä” (*prefabricated patterns, prefabs*, ks. Granger 1998) (Jantunen

2008: 16). Usein tällainen kokeilu ja arvailu johtaa epäkonventionaalisiin valintoihin (mp.), jotka ilmenevät erilaisina natiivikielestä poikkeavina kollokaatioina (ks. Hasselgren 1994; Granger 1998; Howarth 1998; Nesselhauf 2003; 2005). Tarkemmin oppijankielen fraseologisia piirteitä tarkastellaan seuraavassa alaluvussa 3.2.

Tätä tutkimusta merkittävästi koskevat oppijankielen universaalit piirteet ovat epätyypilliset taajuudet (polysemiassa sekä morfologisessa ja semanttisessa primingissa), yksinkertaisuus (muoto- ja merkityselementtien välisen korrelaation voimakkuus) ja kontekstuaalinen epäkonventionaalisuus (semanttinen priming ja paradigmaattinen morfologinen priming).

3.2 Fraseologia oppijankielessä

Koska kieli on fraseologista ja kieltä opitaan fraseologisesti, oppijankielen tutkimuksessa ja opetuksessa tulisi huomioida kielen fraseologinen ulottuvuus. (Granger & Meunier 2008b: 247.) Fraseologia onkin vakiintunut myös oppijankielen tutkimukseen (Granger & Meunier 2008a; Ellis N. 2008: 6–9). Kuten muutakin fraseologista tutkimusta, oppijankielen tutkimusta on tehty erityisesti englannin kielestä korpusten avulla 1980-luvun loppupuolelta alkaen (Granger 2002: 4–5). Korpuksia pidetään ainoana luotettavana lähteenä esimerkiksi taajuuksien tutkimukselle, sillä taajuutta on hyvin vaikeaa havaita intuitiolla, vaikka intuitiolla onkin suuri rooli kielen käytössä. Korpuksilla on mahdollista selvittää mahdollisia ilmauksia, mutta myös niiden todennäköisyyttä. (Granger 2002: 4.)

Fraseologian hallinta on tärkeää, koska fraseologisuuden ymmärtäminen ja kompetenssi muodostaa opittavan kielen mukaisia ilmauksia ovat yhteydessä sujuvaan puhuttuun ja kirjoitettuun kieleen (Howarth 1998: 186; Jantunen & Brunni 2012: 75). Fraseologian hallinta auttaa myös opittavan kielen ymmärtämisessä, kun oppija pystyy hahmottamaan syötöksestä valmiita kokonaisuuksia yksittäisten ilmausten sijaan. Tällöin oppijan ei tarvitse analysoida jokaista ilmausta ja ilmausten keskinäisiä suhteita itsenäisinä yksikköinä. (Vrt. Wray 2002.) Fraseologinen tieto auttaa myös rekisterien ja genrejen hallinnassa, sillä niilläkin on omanlaisensa fraseologiset piirteet. (Jantunen & Brunni 2012: 76.)

Kielenoppijalla on lukuisia mahdollisuuksia päätyä epätyypillisiin rakenteisiin tuottaessaan kohdekielen ilmauksia (Pawley & Syder 1983: 191, ks. Jantunen 2009b), ja lisäksi ongelmia tuottavat morfologiset piirteet juuri fraseologian näkökulmasta. Monet yksittäiset oppijankielen morfologian ja morfosyntaksin piirteet liittyvätkin fraseologisten sääntöjen tunnistamisen, osaamisen ja opettamisen vaikeuteen. Tämä juontaa juurensa siitä, että yleensä oppijalla

ei ole tietoa ilmausten taajimmista (ja siten osaamisen kannalta merkittävimmistä) muodoista tai niiden kontekstuaalisista (esim. kolligationaalisista) ja kontekstuaalisista suhteista. Valenssisäännöt, joiden hallinta perustuu fraseologiasääntöjen havaitsemiseen sekä semanttisten ja tekstin ulkoisten säännösten osaamiseen, ovat tästä hyvä esimerkki. Natiivipuhuja hallitsee monimutkaiset säännöt implisiittisesti, eikä edes kokeneelle opettajalle niiden purkaminen auki ole yksiselitteistä. (Jantunen & Brunni 2012: 78–79.)

Useissa tutkimuksissa on havaittu, että kielenoppijoiden tuottama kieli poikkeaa natiivikielestä nimenomaan fraseologisissa piirteissä (Jantunen & Brunni 2012: 72, ks. Granger 1998; Howarth 1998; Nesselhauf 2003; 2005; Meunier & Granger 2008). Vaikka oppijan tuotos olisi ymmärrettävä ja kieliopillisesti oikein, siinä voi olla epätyypillisiä myötäesiintymiä (ks. Jantunen 2009b: 362). Monilla edistyneilläkin oppijoilla on vaikeuksia natiivinkaltaisen kollokaation ja idiomaattisuuden kanssa, ja siksi monet oppijan tuottamat kieliopillisesti virheettömät virkkeet voivat kuulostaa epäluonnollisilta ja oudoilta (ks. Pawley & Syder 1983; Granger 1998; Howrath 1998). Oppijoiden fraseologian taju on siis natiiveja heikompi, minkä vuoksi oppijat muodostavat natiivien mielestä epätyypillisiä yhdistelmiä (Granger 1998: 152). Tämän vuoksi on perusteltua tarkastella edistyneenkin tason suomenoppijoiden fraseologista osaamista.

Tutkimuksissa fraseologisella piirteellä tarkoitetaan usein kollokaatiota, vaikka fraseologiaan lasketaan myös kieliopilliset ja semanttiset tasot (ks. Granger 1998; Jantunen 2009b; Jantunen & Brunni 2012). Esimerkiksi morfologiset piirteet oppijankielen korpustutkimuksessa ovat jääneet muiden piirteiden varjoon, koska tutkimus on keskittynyt lähinnä englannin ja muiden indoeurooppalaisten vain vähän agglutinoivien kielten tutkimukseen. Morfofonologisten piirteiden tarkastelu taivutukseltaan monipuolisen kielen (kuten suomen) tutkimuksessa on kuitenkin oleellista, koska “fraseologisen yksikön ydin ja kontekstuaaliset elementit suosivat yleensä tiettyjä taivutusmuotoja” (Jantunen & Brunni 2012: 72–73). Ilman useiden erilaisten kieliryhmien tutkimista näkemys fraseologiasta ja sen vaikutuksesta kielenoppimiseen jää vajaaksi (Jantunen 2017: 268), minkä vuoksi suomen kielestä tehdyllä fraseologisella tutkimuksella on antia myös kansainväliselle tutkimuskentälle.

Grangerin ja Meunierin (2008b: 248) mukaan fraseologian osuus toisen ja vieraan kielen opetuksessa on vielä pieni, mutta sillä on valtava potentiaali natiivinkaltaisen sujuvuuden saavuttamisessa. Myös Jantunen (2017: 267) korostaa, että fraseologian ja sen tärkeyden ymmärtäminen kielenoppimisessa ja -opetuksessa on tärkeää, kun tavoitteena on oppia natiivinkaltaista kohdekielen käyttöä. Kuitenkin tutkimuskirjallisuudessa on yksimielinen näkemys siitä, että opittavan kielen fraseologisuus on yksi vaikeimmista asioista omaksua ja opetella (ks.

Granger 1998: 152; Howarth 1998: 186; Granger & Meunier 2008b: 247; Jantunen & Brunni 2012: 75), minkä vuoksi aiheesta tarvitaan lisää tutkimusta.

Fraseologiset ongelmat säilyvät oppijankielessä pitkään, sillä myös edistyneet oppijat tekevät kohdekielen fraseologisista normeista poikkeavia muotoja (Jantunen & Brunni 2012: 76). Esimerkiksi Granger (1998: 158) osoittaa, että edistyneidenkin oppijoiden teksteissä aliedustuvat valmiit kielen rakenteet (*prefabs*) ja yliedustuvat oppijankielen omat rakenteet. Lisäksi tutkimuksessa selviää, että oppijoiden käyttämät kollokaatiot ovat suurimmaksi osaksi keskenään yhteneviä, mikä voi johtua transferin vaikutuksesta¹⁴. Näyttäisi myös olevan niin, että oppijat käyttävät astemääritteitä enemmän rakennuspalikoina kuin osana valmiita rakenteita. Oppijat käyttävät joitain astemääritteitä yleismerkitysyksiköinä, kuten astemääritettä *very*, joka esiintyy merkittävästi ylikäytettynä *-ly*-johtoisten määritteiden kustannuksella. Tulokset kielivät siitä, että oppijoilla ei ole vain heikko, vaan jopa harhaanjohtava keskeisten asioiden havaintokyky. (Granger 1998: 151–152.) Tutkimuksen keskeinen anti on, että edistyneidenkin oppijoiden fraseologiset kyvyt ovat vakavasti rajoittuneet: oppijat käyttävät liian vähän natiivikielelle ominaisia valmiita rakenteita ja liikaa vieraalta kuulostavia (mts. 158).

Tiedetään siis, että kielen fraseologisuus aiheuttaa oppijoille ongelmia. Wray (2002: 200) ehdottaa ilmiölle selitysmallia: että oppijat analysoivat kohdekielen fraseologismeja, mutta vain leksikon tasolla. Tällöin fraseologismien yksityiskohtainen (erityisesti) morfologinen käyttäytyminen, eli *kuinka* sanat yhdistyvät, jää analysoimatta. Kun oppija muodostaa fraseologismia uudestaan, tämä muistaakin vain siinä esiintyneet sanat ja muodostaa ilmauksen sen hetken oppijankielen tasonsa mukaisesti. (Mp.) Toisen ja vieraan kielen heikkoa fraseologista hallintaa selittää siis se, että äidinkieltä opittaessa jo lähtötilanteessa on suuri määrä laajoja ja kompleksisiä ilmauksia, joita ei ole tarvetta purkaa osiin – sen sijaan vierasta kieltä opittaessa aloitetaan pienistä yksiköistä, joita pyritään muodostamaan uudelleen. Fraasit ja lauseet voivat olla osa kielenoppijan kohtaamaa syötettä, mutta niistä huomioidaan vain sanat ja se, kuinka sanat voi “liimata” takaisin yhteen. (Wray 2002: 206.)

Vaikka pidetään yleisesti hyväksyttävänä teoriana, että oppijat eivät kykene kovin tehokkaasti hyödyntämään kohdekielen idiomaattista valintaa (mm. Granger 1998: 146 on todennut, että oppijat hyödyntävät enemmän vapaan valinnan periaatetta), oletusta vastaan on esitetty myös kriittisiä näkemyksiä (ks. Vetchinnikova 2014). Vetchinnikova (2014: 52) esittää, että aiemmat tutkimustulokset eivät välttämättä ole soveltuvia sellaisen päättelyn perustaksi, etteivät kielenoppijat hallitsisi idiomaattista valintaa: tuotosten eroaminen natiivikielestä ei tarkoita

¹⁴ Grangerin tutkimuksessa oppijanenglannin aineisto koostuu yksinomaan ranskankielisistä oppijoista (Granger 1998: 146).

sitä, etteikö oppijankielessä esiintyisi fraseologisia yksiköitä lainkaan. Tähän päätelmään Vetchinnikova (mts. 56) on päätenyt sen vuoksi, että fraseologian tutkimuskentällä fraseologia ymmärretään eri tavoin ja tutkimusmenetelmät ovat vinoutuneet yksipuolisiksi. Lisäksi suurin osa oppijankielen fraseologisesta korpustutkimuksesta pyrkii Verchinnikovin (2014: 49) mukaan osoittamaan todeksi Grangerin (1998: 158) väitteen siitä, että oppijoiden fraseologiset kyvyt ovat rajoittuneet. Oppijankieltä tutkitaankin paljon virheanalyysin avulla, mutta oppijan tuottamia poikkeamia ei välttämättä pidetä silti virheinä. Poikkeamat voidaan tulkita myös oppijankielen ominaispiirteiksi. (Ks. Latomaa 1993: 21; Jantunen 2007.) Usein oppijankielen tutkimuksessa tutkija määrittelee myös etukäteen virhekategorioiden, joiden mukaan oppijankielestä löydetty piirteet luokitellaan. Oppijankieltä on kuitenkin mahdollista tarkastella myös ilman ennako-oletuksia siten, että natiivisuudesta eroavia piirteitä ei pidetä virheinä, vaan poikkeamina ja epätyypillisinä kontekstuaalisina rakenteina, jotka ovat leimallisia oppijankielelle (Jantunen 2007: 69). Näin toimitaan tässä tutkimuksessa.

Vetchinnikova (2014: 216) osoittaa kuitenkin vastoin valtavirtatutkimusta, että kielenoppijat ovat fraseologisesti odotettua taitavampia: oppijat pystyvät siis omaksumaan merkityksikköjä (*units of meaning*) altistuttuaan kohdekielelle, käyttämään niitä kontekstissa ja muistamaan ne idiomaattista valintaa varten. Tutkimuksen ydintulos onkin, että jos kielenoppijan tuottamia rakenteita tarkastelee samalla tavalla kuin natiivin tuottamia, huomataankin, että niiden taustalla olevat prosessointimekanismit ovat huomattavan samankaltaisia (Vetchinnikova 2014: 62).

Tässä tutkimuksessa pyritään etenemään aineiston ehdoilla ilman ennalta määrättyjä odotuksia sen suhteen, kuinka idiomaattista (vs. vapaata) oppijankielen valinta on. Sekä oppijankielen että natiivikielen variantteja tutkitaan itsenäisinä kielimuotoina, joiden eroja tarkastellaan arvottomasti toista paremmaksi kuin toista. Pyrkimys on siis välttää Vetchinnikovan havainto pyrkimyksestä todistaa oppijankielen fraseologian heikkoutta. Sen sijaan tavoite on selvittää, millaista oppijankielen fraseologia on (toki kuitenkin verraten sitä prestiisiin natiivikielen). Tutkimuksen lähtöoletus on siis, että oppijankielessä on omanlaisia fraseologisia piirteitä, jotka erottavat sen natiivikielestä.

3.3 Morfologinen ja semanttinen priming oppijansuomessa

Kielenoppijat eivät hallitse morfologista primingia, kuten muitakaan fraseologisia piirteitä, samalla tavalla kuin natiivipuhujat. Tämä paljastuu tarkastelemalla erilaisten taivutusmuotojen frekvenssejä, jotka oppijankielessä eroavat natiivipuhujien tuottamista muodoista. Jantusen ja Brunnin (2012: 94) tutkimustulosten mukaan ”oppijankielen fraseologisen yksikön ytimen paradigmaattinen valinta” eroaa kohdekielestä. Lisäksi fraseologisen yksikön (tässä tapauksessa ytimen ja sen kollokaattien muodostaman kokonaisuuden) morfologinen priming eroaa oppijan- ja kohdekielessä. (Mts. 94.)

Syynä kielenoppijoiden tuottamaan epätyypilliseen morfologiseen primingiin on pyrkimys välttää tiettyjen muotojen käyttöä. Jantusen ja Brunnin (2012: 94) tutkimuksessa tämä näkyy esimerkiksi nominin IHMINEN käytössä monimutkaisen passiivirakenteen asemesta (esimerkiksi *kirkon jälkeen ihmiset menevät kotiin syömään* passiivisen *kirkon jälkeen mennään kotiin syömään* sijaan). Yleisluonteinen IHMINEN onkin nollapersoonan ohella yksi suomenoppijoiden strategisista keinoista (Jantusen ja Brunnin 2012: 82 mukaan Seilonen 2011). Toinen syy epätyypilliseen primingiin voi löytyä rakenteellisesti hyvänmuotoisista ilmauksista, jotka ovat kuitenkin natiivikielessä harvinaisia. Esimerkiksi oppijansuomessa esiintyvä ilmaus ’perheessä on # ihmistä’ on rakenteellisesti oikein, mutta natiivisuomessa ilmausta ei käytetä, koska tyypillisempää on luetella perheenjäsenet suoraan. (Jantunen & Brunni 2012: 85–86.)

Semanttista primingia on oppijankielestä tutkittu toistaiseksi melko vähän. Kuitenkin Jantunen (2017) tarkastelee myös semanttista primingia *kello*-sanan holistisessa tutkimuksessaan. Jantunen (2017: 266) osoittaa, että *kello*-sanin semanttisilla primingeilla on eroa oppijankielen ja natiivikielen aineistoissa: esimerkiksi oppijankielessä *kello* esiintyy nukkumista (semanttinen joukko ‘sleeping’) merkitsevien sanojen kanssa useammin kuin natiiviaineistossa ja natiiviaineistossa katsomista (semanttinen joukko ‘looking at’) merkitsevien sanojen kanssa useammin kuin oppijankielen aineistossa. Molemmissa aineistoissa *kello* esiintyy taajimmin ajanilmauksissa, mutta oppijankielessä tämä korostuu natiivikieleen verrattuna.

Vaikka morfologisen ja semanttisen primingin tutkimusta oppijansuomesta on siis verrattain vähän, niiden tiedetään eroavan natiivisuomesta. Oppijankielessä käytettyihin muotoihin voi tosin vaikuttaa myös konkreettiset ja opetuksessa painotetut rakenteet, jolloin variaatio jää natiivipuhujia pienemmäksi (Jantunen & Brunni 2012: 94.)

4 TUTKIMUSAINEISTOT

4.1 Korpukset tutkimusaineistona

Korpukset ovat laajoja ja järjestelmällisesti kerättyjä kokoelmia luonnollisesta kielestä (Tieteen termipankki s. v. *corpus*). Korpustutkimuksessa formaatti on merkittävä tekijä. Korpuksiin voi-kin koodata missä vain muodossa tuotettua kieltä: on olemassa sekä puhekielen että kirjoitetun kielen korpuksia (McEnery & Hardie 2014: 3). Lisäksi jotkin korpukset rekisteröivät paralingvistisiä piirteitä, kuten eleitä (McEneryn & Hardien 2014: 3 mukaan Knight, Evans, Carter & Adolphs 2009b). Myös viittomakielen korpuksia on olemassa (McEneryn & Hardien 2014: 3). Tässä tutkimuksessa käytetään kirjoitetun kielen aineistoja. Kirjoitettu oppijankieli on kenties harkitumpaa tuotosta kuin puhe, jolloin oppijalla on ollut aikaa hioa ilmaisuun. Tämä palvelee tutkimusasetelmaa, kun tarkoitus on selvittää verbin käytön systemaattisuuksia: puhetilanteissa ei ole yhtä paljon aikaa miettiä tuotoksen muotoa. Toisaalta puhekielen fraseologian tutkimus olisi aivan yhtä tärkeää, mutta se on rajattu pois tästä tutkimuksesta.

Tässä kohti onkin paikallaan huomauttaa, että aineiston formaatin vuoksi tutkimuksen tulokset pätevät vain kirjoitettuun suomen kieleen. Tämä on tärkeää siksi, että formaatit itsessään luovat kielitieteellisesti merkittäviä eroja. Laajamittaisten puhutun ja kirjoitetun kielen erojen tutkimusten tuloksena on ymmärretty, kuinka merkittävästi erilaisia puhuttu ja kirjoitettu kieli ovat kieliopin tasolla. On myös esitetty, että puhutun ja kirjoitetun kielen kieliopit eivät vain eroaisi toisistaan, vaan että ne olisivat toisistaan täysin erilliset (ks. Brazil 1995). Myös suomen kielessä kirjoitettu ja puhuttu kieli eroavat toisistaan merkittävästi kieliopin osalta (ks. Hakulinen 2003; Viinikka & Voutilainen 2013). Tämän vuoksi kielitieteelliseltä kannalta on merkittävää, mitä formaattia aineisto on. Kyse ei ole siis vain aineistonkeruutavasta tai teknisistä asioista. (McEnery & Hardie 2014: 5.)

Aineiston tyyppin valinnalla on myös väliä korpustutkimuksessa, koska luonnollisesti aineiston pitäisi vastata tutkimuskysymyksiin. Tästä herää kysymys: Kuinka voidaan varmistaa, että vastaavuus on tarpeeksi hyvä? Sopivan aineiston valintaan on kehkeytynyt kaksi yleistä menettelytapaa. (McEnery & Hardie 2014: 6.) Näistä ensimmäinen on tarkkailukorpus (*monitor corpus*) (Sinclair 1991: 24–26), joka laajenee jatkuvasti ja sisältää ajan myötä yhä enemmän tekstejä. Toinen menettelytapa on nimeltään näytekorpus (*sample corpus*) (McEneryn & Hardien 2014: 6). Näytekorpus kootaan harkiten noudattaen tarkasti rajattua otantakehikkoa (*sampling frame*), ja se kuvastaa kieltä niin kuin se esiintyy tietyssä ajankohdassa (McEnery 2014:

6). Otantakehikko määrittelee korpuksen kielen tyyppin sekä perusjoukon (*population*), jota tutkimuksella pyritään luonnehtimaan. Perusjoukko on teoreettinen tila otannan kohteena olevassa kielessä. (McEnery & Hardie 2014: 8–9.) Tässä tutkimuksessa oppijansuomen korpus tarjoaa siis teoreettisen tilan oppijankielestä ja natiivisuomen korpus vastaavasti natiivikielestä. Näytekorpuksen koostamisessa on tärkeää varmistaa, että kerätty data on jollain tavalla tyyppillistä, jotta aineistoon ei aiheutuisi vinoumia. Näytekorpuksissa ajatuksena on siis luoda mahdollisimman hyvin todellisuutta kuvaava aineisto. (McEnery & Hardie 2014: 9.) Molemmat tutkimuksessa hyödynnetyt korpuksot edustavat selkeästi näytekorpuksia, sillä ne on kerätty tarkkojen raamien puitteissa edustamaan spesifiä kielimuotoa.

Korpusaineistot eroavat toisistaan myös sen osalta, onko aineistoon koodattu tietoa¹⁵. Tällaista koodausta kutsutaan annotaatioksi (*annotation*). (McEnery & Hardie 2014: 13.) Korpuksot voidaan annotoida kielitieteellisellä tiedolla siten, että voi myös tutkia jotain muuta kuin varsinaista kielellistä aineista, esimerkiksi jotain kieliopillista sijaa tai verbin muotoa. (Mts. 29–30; esimerkit omia.) On kuitenkin tärkeää huomata, että annotaatio ei välttämättä ole johdonmukaista (McEnery & Hardie 2014: 32). Johdonmukaisuusasiaa voi käsitellä myös korpusten välillä, koska toiset annotoinnit ovat tarkempia kuin toiset. Annotointimalleissa voi myös olla eroja ja annotointia voi tehdä erilaisilla ohjelmilla, jotka annotoivat hieman eri tavalla. Eri korpusten välinen epäjohdonmukaisuus annotoinnissa voi muodostua ongelmaksi, jos tutkii useampaa annotoitua korpusta samaan aikaan. (McEnery & Hardie 2014: 33.) Annotaatio kuitenkin puolustaa paikkaansa tieteen toistettavuuden näkökulmasta: löydökset on helpompi toistaa uudestaan, kun annotaatiot ovat kaikille saatavilla (mts. 32–33).

Tässä tutkimuksessa aineistot on annotoitu eri tavoin, mutta sen ei koeta haittaavan tutkimusta, koska molemmista aineistoista on hyödynnety lemmatisointia, joka nähdään vähemmän altis virheille kuin esimerkiksi syntaktinen annotaatio. Lemmatisoinnin suhteen ei myöskään tullut vastaan ongelmia.

4.2 Kansainvälinen oppijansuomen korpus (ICLFI) ja Käännössuomen korpus (KSK)

Tämän tutkimuksen oppijankielen aineisto on kerätty **Kansainvälisestä oppijansuomen korpuksesta** (ICLFI) (Jantunen 2011b). ICLFI on kirjoitetun kielen oppijansuomen korpus, jonka kokoaminen alkoi vuonna 2007. Korpus sai alkunsa Oulun yliopiston tutkimusprojektissa

¹⁵ Osa tutkijoista vastustaa annotointia sen lukuisista hyödyistä huolimatta, koska sen nähdään sotkevan kieliaineksen “puhtauden” (McEnery & Hardie 2014: 14).

Korpustutkimus oppijankielen kielikohtaisista ja universaaleista ominaisuuksista, jota tehtiin yhteistyössä viiden yliopiston kanssa. (Jantunen 2011b: 86.) ICLFI sisältää n. 1,0 miljoonaa sanetta 22 eri äidinkieltä edustavilta suomi vieraana kielenä -opiskelijoilta. Korpus sisältää yhteensä 5400 tekstiä. (Jantunen 2011b: 89.)

Tarkemmin ICLFI on vieraan kielen korpus (vrt. toinen kieli), sillä aineisto on kerätty ulkomaisissa yliopistoissa (Jantunen 2011b: 92). Näiden yliopistojen suomen kielen opettajat ovat keränneet opiskelijoiden kirjalliset tehtävät talteen, ja korpukseen on valikoitu tekstinäytteitä opettajien laatimista koosteista. Koska tekstit on kerätty sisällöltään, painopisteiltään ja vaikeustasoiltaan eroavilta kursseilta, korpus sisältää tekstejä erilaisista genreistä ja tekstityypeistä. Suurin osa teksteistä on kuvailevia esseitä, mutta korpus sisältää myös kertomuksia, kuvauksia, päiväkirjoja, arvosteluja, referaatteja, mielipidekirjoituksia, vastineita, uutisia, sähköpostiviestejä, kirjeitä, satiireja ja työhakemuksia. (Spoelman 2013: 173; Jantunen & Pirkola 2015: 92.) ICLFI on siis monitekstilajinen korpus. Usea tekstilaji takaa sen, että tulokset eivät ole tietyn genren ominaispiirteitä, vaan todella oppijankielen. (Jantunen 2011b: 89–90.)

Tutkimuksessa hyödynnetään ICLFI:n aineistoa eurooppalaisen viitekehyksen (EVK 2003: 47; Opetushallitus 2003) taitotasojen B2–C2 väliltä siten, että jokainen lähtökieli on aineistossa mukana. Taitotasot ovat asteikon yläpäässä, ja taitotasokuvauksen mukaan kyseisillä taitotasoilla oppija selviää säännöllisessä vuorovaikutuksessa syntyperäisten kanssa (B2) ja selviää monissa vaativissa kielenkäytön tilanteissa (C1–C2). Tutkimuksessa tarkastellaan edistyneiden kielenoppijoiden tekstejä, koska fraseologiset virheet rasittavat myös edistyneitä oppijoita (Jantunen 2009b: 361). Sujuvasta kielenkäytöstä on tarkoituksenmukaisempaa tutkia systemaattista kielellisen ilmauksen käyttöä, kun alkeistason virheet on muuttujina rajattu pois.

Tämän tutkielman natiiviaineisto on kerätty **Käännösuomen korpuksesta (KSK)** (Jantunen 2004: 5 mukaan Mauranen 2000). KSK on koottu Joensuun yliopistossa *Translated Finnish and Translation Universals* -projektissa vuosien 1998–2000 aikana. Käännösuomen korpusten alkuperäissuomen aineistossa, jota käytän natiivisuomen aineistona, on noin 3,8 miljoonaa sanetta. Koko korpuksessa saneita on lähes 10 miljoonaa. (Jantunen 2004: 40; Spoelman 2013: 186).

Nimestään huolimatta Käännösuomen korpus sisältää muutakin kuin käännettyä kieltä, alun perin suomeksi kirjoitettua kieltä (Spoelman 2013: 186). Natiivisuomen osakorpusta kutsutaan alkuperäissuomen aineistoksi. Tässä osakorpuksessa on ICLFI:n tapaan useita genrejä: tieteellisiä tekstejä, populaareja tieteellisiä tekstejä, aikuisille suunnattua fiktiota (kertomakirjallisuutta, salapoliisi- ja viihdekirjallisuutta), aikuisille suunnattua ei-fiktiota (muistelmia) ja lapsille suunnattua fiktiota (Jantunen 2004: 40; Spoelman 2013: 187).

4.3 Aineistojen vertailukelpoisuus

ICLFI:lle ei ole koottu samanaikaisesti natiivisuomen vertailukorpusta. Tämän vuoksi täysin vastaavaa aineistoa ei ole, ja täytyy tyytyä tarpeeksi samankaltaiseen vertailuaineistoon. Tässä tutkimuksessa on päädytty Käännössuomen korpuksen alkuperäissuomen osakorpukseen, koska se sisältää laajasti erilaisia genrejä, kuten myös ICLFI. Lisäksi alkuperäissuomen osakorpuksen tekstit ovat ammattilaisten kirjoittamia, joten sen voidaan ajatella edustavan standardia suomen kirjoitettua yleiskieltä. Samaa aineistoparia ovat käyttäneet tutkimuksissaan muun muassa Jantunen ja Brunni (2012), Spoelman (2013) ja Jantunen (2017).

Asetelmaa on myös kritisoitu sillä perusteella, että oppijankielen tekstien vertaaminen ammattitason natiiviteksteihin nähdään epärealistisena (ks. Lorenz 1999: 71–73). Kriittisen näkemyksen mukaan olisi parempi, että oppijankieltä verrattaisiin natiiviopiskelijoiden teksteihin kuin ammattilaisten teksteihin, mikäli tavoitteena on vertailla, analysoida ja kuvata oppijankieltä mahdollisimman oikeudenmukaisesti (Hyland & Milton 1997: 184 Spoelmanin 2013: 185 mukaan).

Myös tätä näkökulmaa on kritisoitu, sillä natiivioppijat tekevät erilaisia ortografisia virheitä (Spoelmanin 2013: 185 mukaan Cutting 2000), minkä vuoksi ammattilaisten tekstit tarjoaisivat paremman vertailukohdan: nämä tekstit edustavat kielimuotoa, jota oppijat yrittävät tavoitella (Spoelmanin 2013: 185 mukaan Ädel 2006: 206–207). Kohdekielen natiivipuhujien tuottamasta tekstistä koostuvaa korpusta pidetäänkin standardinmukaisena kielen normina, jota vasten voi tarkastella oppijankielen piirteitä (Leech 1998: xv). Tässä tutkimuksessa päädyttiin ammattilaisten tuottamaan natiivikielen vertailuaineistoon, alkuperäissuomen osakorpukseen, myös sen vuoksi, että korpusparilla on tehty tutkimuksia runsaasti aiemminkin. Tästä johtuen tutkimustulokset ovat paremmin vertailtavissa aiempaan tutkimukseen. Oppijankielen ja natiivikielen vertailuasetelman käsittelyä jatketaan seuraavaksi menetelmällisesti.

5 TUTKIMUSMETODIT

5.1 Kontrastiivinen oppijankielen analyysi (CIA)

Oppijankielen tutkimuksessa on mielekästä tehdä kahdenlaista vertailua (Granger 2015: 8):

1. oppijankielen vertailua natiivikielen, joka nähdään kielenoppimisen tavoitteena
2. tietyyntyyppisen oppijankielen variantin vertailua toisenlaiseen, esimerkiksi eri äidinkielisten oppijankieltä.

Menetelmää, jossa edellä mainittuja vertailuja tehdään, kutsutaan kontrastiiviseksi oppijankielen (välikielen) analyysiksi (CIA; *Contrastive Interlanguage Analysis*) (ks. Granger 2015). CIA-metodia on luonnehdittu myös virheanalyysin sekä oppijankielen ja natiivikielen vertailun yhdistäväksi metodiksi (Jantunen & Brunni 2012: 72).

Ennen Grangerin (1996) esittelemää CIA-mallia Péry-Woodley (1990) oli käyttänyt oppijankielen ja natiivikielen vertaamisen menetelmästä nimitystä *comparative interlanguage research*, joka tosin eroaa CIA-mallista siten, että CIA sisältää myös erilaisten oppijankielten varianttien vertailun (Granger 2015: 9). Tässä tutkimuksessa ollaan siis hyvin lähellä *comparative interlanguage research* -metodia, mutta tätä maisterintutkielmaa ja tyypillisesti CIA-menetelmää yhdistävän oppijankielen korpustutkimuksen vuoksi tutkimus halutaan liittää metodologisesti CIA-metodiin, jolloin myös vertailu muiden vertailevien oppijankielen korpustutkimusten kanssa on helpompaa.

Vertailevan tutkimuksen tyypillinen lähtökohta on jukstapositioneja kaksi tai useampi erillistä asiaa (tässä tapauksessa oppijansuomi ja natiivisuomi), jotta niiden väliltä olisi mahdollista paikantaa samankaltaisuuksia ja eroavaisuuksia. Tällöin on mahdollista paljastaa sellaisia aspekteja, joita ei olisi voinut paljastaa, mikäli asiat olisivat olleet eristyksissä (Granger 2015: 7 mukaan Strawn 2009: 117). Grangerin (2015: 7) mukaan vertaileva tutkimusote onkin luonnollinen tapa “saada oppijankielen korpukset puhumaan”.

CIA:n näkemys sanastosta on fraseologisen viitekehyksen kanssa linjassa: menetelmässä on yksittäisten sanojen sijaan keskitytty useasanaisiin yksiköihin. CIA-menetelmän puitteissa onkin tutkittu muun muassa kollokaatioita ja kolligaatioita, jotka ovat merkittävästi muutaneet käsitystämme oppijankielen fraasien ja monisanaisten yksiköiden varastosta, “fraasistosta” (*phrasicon*, ks. Paquot & Granger 2012). (Granger 2015: 10.)

CIA-menetelmä on kehitetty edistyneen tason kielen omaksumisen tutkimiseen, sillä SLA-tutkimus on aiemmin keskittynyt enemmän alkeistason tutkimiseen. Tämän vuoksi

edistyneestä oppijankielestä ei ole ollut tarpeeksi tietoa, eikä vastaavasti edistyneiden oppijoiden tarpeisiin ole ollut soveltuvaa pedagogista materiaalia. CIA-menetelmän avulla on saatu parempi käsitys sanastollis-kieliopillisista ja diskursseihin liittyvistä ominaisuuksista, jotka luonnehtivat edistynyttä oppijankieltä. (Granger 2015: 11.) Kuten tutkielmassa on jo aiemmin tullut ilmi, tässäkin tutkimuksessa CIA-metodia sovelletaan nimenomaan edistyneeseen oppijansuomeen: EVK:n taitotasoihin B2–C2.

CIA-metodia on käytetty monenlaisissa tutkimuksissa: Learner Corpus Association -nettisivuston¹⁶ mukaan metodia on käytetty ainakin osittain yli 1000 tutkimuksessa (Granger 2015: 9). Vaikka CIA-metodi aluksi keskittyi englannin kielen L2-tutkimukseen, se on levinnyt myös muihin kieliin, kuten suomen kieleen (mm. Jantunen & Brunni 2012). Jantusen ja Brunnin (2012: 72) mukaan CIA-menetelmä ei ole kuitenkaan rantautunut sellaisenaan oppijansuomen tutkimukseen. Muun muassa tutkimuskysymykset ovat osittain erilaisia suomalaisessa oppijankielen tutkimuksessa, sillä esimerkiksi virheanalyysi ei ole oppijansuomen korpustutkimuksessa samalla tavalla keskiössä kuin englanninkielisessä korpustutkimuksessa. Virheanalyysia ei tehdä tässäkin tutkimuksessa, vaan oppijan- ja natiivikielen fraseologiaa piirteitä vertailaan arvottamatta oppijankielen poikkeamia natiivikielestä virheiksi.

CIA-metodin kahdesta haarasta suositumpi on oppijankielen vertailu natiivikieleen. Vaikka tämän tyylinen vertailu nähdään myös kiistanalaisena, on ymmärrettävää, että tutkijat ovat omaksuneet sen edistyneen tason kielen omaksumisen tutkimisen jatkumoon. On järkevämpää verrata edistyneitä oppijoita natiiveihin kuin tehdä niin alkeistason oppijoiden kanssa, koska edistyneiden oppijoiden ja natiivien kielitaidon tasojen väli on pienempi ja vertailu on tehokas heuristinen työkalu sen supistamiseksi edelleen. Kielen piirteiden yli- ja alikäyttö on vaivatonta tunnistaa sopivilla työkaluilla ja metodeilla, ja nämä voivat antaa sysäyksen jatkotutkimuksille. Sen sijaan kielen piirteiden virheellinen käyttö on paljon vaikeampaa huomata tämän tyyppisellä vertailulla, ja siksi sitä yleensä tutkitaankin tietokoneavusteisella virheanalyysillä mieluummin kuin CIA-menetelmällä. Tästä huolimatta vertailu oppijankielen ja natiivikielen välillä voi auttaa paljastamaan joitakin virheitä (tai epätyypillisyyksiä), sillä ylikäyttö on usein merkki virheestä. Esimerkiksi ranskankieliset englannin oppijat ylikäyttävät ilmausta *on the contrary*, koska sillä on virheellinen yksi yhteen -vastaavuus ranskan *au contraire* -konnektorin kanssa. (Granger 2015: 11.)

CIA-metodia on myös kritisoitu oppijankielen ja natiivikielen välisestä vertailuasetelmasta. Kritiikki kohdistuu vertailuasetelman virheellisyyteen (*comparative fallacy*) ja normin

¹⁶ <https://www.learnercorpusassociation.org/>

määrittelyn ongelmallisuuteen. (Granger 2015: 13.) Vertailuasetelman virheellisyyttä on kritisoitu siten, että oppijankieltä pitäisi tutkia yksinään eikä suhteessa natiivikieleen, koska tällöin oppijankieli nähdään väkisin puutteellisena (ks. Bley-Vroman 1983: 2; Selinker 2014: 230; Larsen-Freeman 2014: 217). CIA-menetelmää on kritisoitu myös siitä, että se olettaa oppijoiden tavoittelevan natiivipuhujan normeja (Hunston 2002: 211).

Granger (2015: 13) vastaa kritiikkiin lainaamalla Tenjfhordia, Hagenia ja Johansenia. (2006: 101–102) siitä, ettei teoriaa ja menetelmää pidä sekoittaa keskenään:

the TL [target language] description imposed upon the text by the analyst provides an objective method for pointing out phenomena that any adequate theory of SLA must be able to account for on a theoretical level

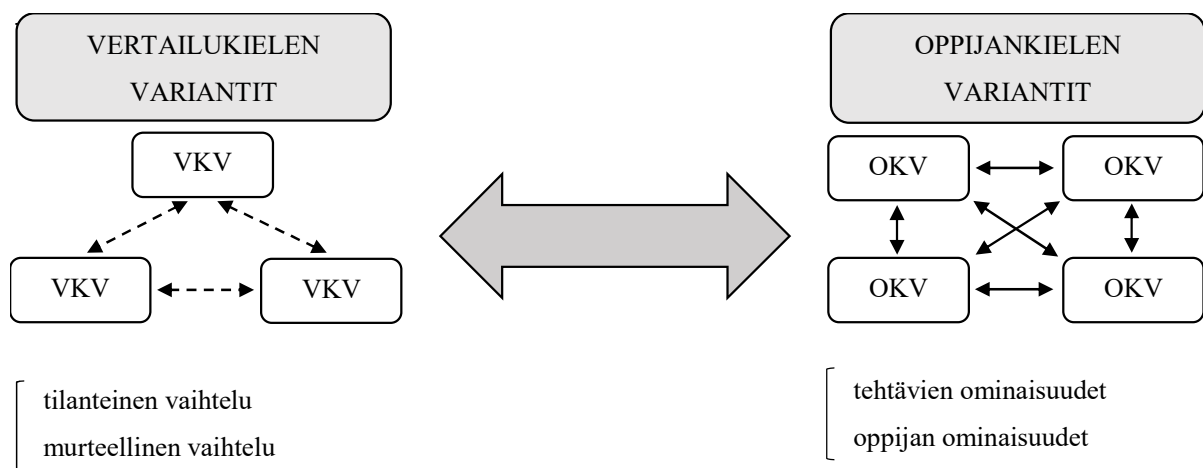
Lisäksi Granger (2015: 13) vetoaa Sung Parkin (2004: 3) huomioon, että jokaisen SLA-tutkimuksen käsitys oppijankielestä pitää implisiittisesti sisällään kohdekielen vaanimassa taustalla. Kaikki tutkimukset, jotka vertailevat oppijoiden taitotasoa, perustuvat tosiasiallisesti taustalla olevaan kohdekielen normiin, sillä taitotasoa arvioidaan yleensä kohdekieli mielessä pitäen (Granger 2015: 14). On perusteltua määrittää, missä oppijankieleessä on todennäköisimmät poikkeamat natiivikielestä ja missä määrin ne esiintyvät. Voisi jopa todeta, että on tutkijoiden työtä pyrkiä selvittämään ainakin osa niistä oppijankielen alasysteemeistä, jotka perustuvat kohdekielen kieliopin lisäksi natiivien kielisysteemeihin ja universaalien kieliopin periaatteisiin. (Grangerin 2015: 14 mukaan Lardiere 2003: 140.) White (Grangerin 2015: 14 mukaan 2003: 27) lisää, että kyse ei ole siitä, etteikö oppijankieltä ja natiivikieltä saisi koskaan verrata keskenään – ongelmallista on, kun tietynlaisia johtopäätöksiä vedetään, jos oppija ei kykene suoriutumaan tismalleen kuten natiivit.

Asetelman kritisointi on johtanut myös termien yli- ja alikäyttö (*over- and underuse*) kyseenalaistamiseen. Sanakirjoihin nojaten kritiikki on perusteltua, sillä esimerkiksi ylikäyttö tarkoittaa, että jotakin käyttää liikaa tai liian usein (Granger 2015: 19). Oppijankorpustutkimusta koskevassa kirjallisuudessa kuitenkin selviää, että termit ovat neutraaleja (ks. Gilquin & Paquot 2008: 38). Termit ovat kuvaavia, eivät ohjailevia (Granger 2015: 18). Granger (mts. 19) kuitenkin myöntää jälkikäteen, että kenties termit yli- ja aliedustus (*over- and underrepresentation*) olisivat kuvanneet ilmiötä paremmin. Vaikka Grangerista (2015: 19) jo vakiintuneita termejä yli- ja alikäyttö voi edelleen käyttää, tässä tutkimuksessa käytetään neutraalimpia käsitteitä yli- ja aliedustus, jolloin ilmiöt kytkeytyvät virheiden sijaan osaksi oppijankielen piirteiksi (ks. Jantunen 2008).

Grangerin (2015: 14) mukaan CIA-vertailuasetelmaa ei kuitenkaan yksimielisesti pidetä virheellisenä SLA-yhteisössä: kritiikki ei ole ollut tarpeeksi vakuuttavaa, jotta oppijankielen

corpustutkijat olisivat hylänneet oppijankielen ja natiivikielen välisen tutkimusasetelman. Kriittikki on kuitenkin terve muistutus siitä, että oppijankieltä pitäisi tutkia myös itsenäisenä kielimuotona. Grangerin (mp.) mukaan oppijankielen ja natiivikielen vertailu on kuitenkin pedagogisesta näkökulmasta mielekästä, sillä vertailun tulokset osoittavat, mitä oppijat tekevät samalla tai eri tavalla tai osittain samalla tai eri tavalla kuin natiivit jollakin tietyllä osa-alueella. Tuloksia voi hyödyntää laajasti pedagogisissa sovelluksissa. Tässä tutkimuksessa ei oteta kantaa oppijankielen oikeellisuuteen tai virheellisyyteen, vaan ilmiöitä tarkastellaan oppijankieleen kuuluvina piirteinä, samanlaisina tai erilaisina suhteessa natiivikielen.

Tässä tutkimuksessa CIA:ta, tai tarkemmin sen päivitettyä (Granger 2015) versiota (ks. kuvio 3) sovelletaan sen ensimmäisen vertailuasetelman osalta eli vertailukielen varianttien (tässä tapauksessa natiivisuomen varianttien) ja oppijankielen varianttien (tässä tapauksessa oppijansuomen varianttien) vertailumenetelmänä: kielimuodoista vertaillaan samoja kielen piirteitä (tässä tapauksessa SAADA-verbin fraseologisia piirteitä) keskenään.



Kuvio 3. Suomeksi mukailtu malli uudistetusta CIA-menetelmästä¹⁷.

Kuviossa 3 havainnollistetaan vertailuasetelmaa tarkemmin. Kyseessä ei siis ole yhden natiivikielen variantin ja yhden oppijankielen variantin vertailu keskenään, vaan tässäkin tutkimuksessa käytetyt käsitteet natiivisuomi ja oppijansuomi koostuvat useista eri varianteista. Granger (2015: 17–18) huomauttaa, että vertailukielen käsite tekee selväksi sen, että korpuksen ei välttämättä tarvitse edustaa kielen normia, vaan mielekkäämpää voi olla vertailla esimerkiksi oppijankielen argumentatiivisia esseitä natiiviopiskelijoiden argumentatiivisia esseitä sisältäviin korpuksiin kuin yleiskorpuksiin. Asetelma painottaa sitä, ettei vertailukorpuksen ole

¹⁷ Kuvio 3 on mukailtu Grangerin (2015: 17) mallista.

tarkoitukseen edustaa kielen normia. Tästä syystä useita genrejä sisältävän ICLFI-korpuksen vertailukorpuksena on vastaavasti useita genrejä edustava Käännössuomen korpuksen alkusuomen aineisto, jonka ei myöskään ole tarkoitus edustaa suomen kielen normia. Vaikka tässä tutkimuksessa ei vertailla eri oppijankielen variantteja keskenään, on hyvä huomata, että ICLFI:stä kerätty oppijansuomen aineisto sisältää useita eri lähtökieliä, erilaisia tehtäviä ja eri lähtökohdista tulleita oppijoita, sillä luotettavien johtopäätösten tekeminen olisi vaikeaa vain yhden oppijankielen variantin perusteella (Granger 2015: 12).

5.2 Korpuspohjainen tutkimusmenetelmä

Korpus tutkimus ei ole niinkään kielentutkimuksen osa-alue, vaan pikemminkin sen väline ja metodi. On tärkeää muistaa, että korpuslingvistiikka on heterogeeninen osa-alue, vaikka sen piirteet ovat osin samankaltaisia eri suuntauksissa. Korpusmenetelmillä on mahdollista tutkia laajasti kielen eri osa-alueita, ja sillä voi tutkia sellaisia asioita, joita olisi lähes mahdotonta tutkia ilman korpuksia: korpusaineistot ovat hyvin kookkaita, joten niiden tutkiminen käsin olisi turhan työlästä. (McEnery & Hardie 2014: 1–2.)

Tässäkin tutkimuksessa käsiteltävät aineistot on kerätty laajoista miljoonien saneiden korpuksista. Korpuksen hyödyn nähdäänkin piilevän niiden koossa – laajoista aineistoista voi löytää kielestä yllättäviä systemaattisuuksia, joita ei huomaisi pienemmästä aineistosta. Tämän vuoksi korpus tutkimus sopii erityisen hyvin fraseologian, piilevien rakenteiden, tutkimiseen (ks. Jantunen 2009b: 373). Tässä tutkimuksessa tarkasteltuja SAADA-verbin eri merkityksiä, semanttista ja morfologista primingia olisi hyvin haasteellista tutkia ilman korpuksista saatavaa tietoa.

McEnery ja Hardie (2014: 1) määrittelevät korpuslingvistiikan seuraavasti:

[D]ealing with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions.

Korpus tutkimuksessa tekstillä tarkoitetaan dataa, jota kone voi lukea (kirjoitettua tekstiä, mutta myös esim. kuvaa) (McEnery & Hardie 2014: 2). Vapaasti suomennettuna korpuslingvistiikka siis on sellaisen koneella luettavan tekstijoukon käsittelyä, mitä pidetään tarkoituksenmukaisena lähtökohtana spesifien tutkimuskysymysten tutkimiselle.

Kenties merkittävin korpus tutkimuksia erotteleva ominaisuus on lähestymistapa aineistoon: onko kyse korpuspohjaisuudesta vai korpusvetoisuudesta. Korpusvetoinen lingvistiikka

nojaa aineistoon jo aivan tutkimuksen alkuvaiheessa. Tässä vaiheessa tutkijalla on kielellisten elementtien asemesta mielessä laaja tutkimuskehys, mutta pienemmät tutkittavat yksiköt valikoituvat aineistosta. (Jantunen 2011a: 50.) McEneryn ja Hardien (2014: 6) mukaan korpusvetoinen lingvistiikka kiistää korpuslingvistiikan luonnehdinnan metodiksi ja se pitääkin itseään ainoana oikeana lähteenä kieleen liittyvän hypoteesin muodostamiselle (ks. Tognini-Bonelli 2001: 84). McEneryn ja Hardien (2014: 6) mukaan kaikkea korpuslingvistiikkaa voi kuitenkin kuvata korpuspohjaiseksi, ja he kiistävät ajatuksen korpusten itsenäisestä teoreettisesta arvosta. Tästä seikasta tässä tutkimuksessa ollaan heidän kanssaan samaa mieltä, sillä korpuksetkin on aina koottu jollain periaatteella eivätkä ne siten objektiivisesti voi esittää koko kielen kirjoa (ks. Jantunen 2009b: 375).

Korpuspohjaisuus sen sijaan tarkoittaa sitä, että korpusdataa käytetään jonkin teorian tai hypoteesin todistamiseksi, kumoamiseksi tai jalostamiseksi. Korpuspohjaisessa tutkimuksessa hypoteesi on usein nousseet muualta kuin aineistosta. (McEnery & Hardie 2014: 6; Tognini-Bonelli 2001: 65; Jantunen 2011a: 49.) Korpusvetoisuutta käytetään siis tutkimaan aineistosta nousevia ilmiöitä, kun taas korpuspohjaisuutta käytetään tutkimaan ennalta valittua ilmiötä aineiston avulla. Tämä tutkimus noudattaa nimenomaan korpuspohjaista lähestymistapaa: tutkittavat kielen elementit on valittu etukäteen ja tarkoitus on tarkentaa ja jalostaa oppijankielen fraseologian kuvausta.

Korpuslingvistiikassa kvantitatiivinen ja kvalitatiivinen ote ovat yhtä tärkeitä. Tosin tutkimustulosten tärkeys riippuu osin siitä, miten hyvin valittu korpusaineisto vastaa valittuun tutkimuskysymykseen. Tämä on jossain määrin ilmiselvää: korpusta käytetään vastaamaan tutkimuskysymykseen, johon se on parhaiten laadittu vastaamaan. Äärimmäisenä esimerkkinä swahilin kielen luokittelusysteemiä ei kannata tutkia englanninkielisistä sanomalehdistä kootusta korpuksesta. Myös korpusaineiston homo- tai heterogeenisyys määrittelee, onko aineisto sopiva tietylle tutkimuskysymykselle. (McEnery & Hardie 2014: 2.) Koska tässä tutkimuksessa tarkastellaan oppijansuomen fraseologisia piirteitä ja verrataan niitä natiivisuomen piirteisiin, oppijansuomen korpus sekä natiivisuomen korpus ovat tarkoituksenmukaisesti valittuja tutkimusaineistoja (aineistojen vertailukelpoisuudesta ks. luku 4.3).

McEneryn ja Hardien (2014: 14) mukaan Leech (1992) on esittänyt, että korpustutkimuksen merkittävimpiä hyötyjä on se, että tutkijat voivat lähestyä kieltä (luonnon)tieteellisin metodein. Leechin (mts. 112) mukaan ydinperiaate tässä lähestymistavassa on täysi vastuunalaisuus (*total accountability*). Jos korpusta lähestyy tietty teoria mielessään, käy helposti niin, että vahingossa aineistosta poimii vain kyseistä teoriaa tukevia esimerkkejä (tätä kutsutaan vahvistusharhaksi, *confirmation bias*). Tällä tavalla teoriaa on mahdotonta todistaa vääräksi.

(McEnery & Hardie 2014: 14–15.) Täyden vastuunalaisuuden periaatteen mukaan tutkija ei saa valikoida teorialle suotuisaa osajoukkoa aineistosta, mikä on korpuspohjaisen tutkimuksen vaara. Kun korpusta käyttää hypoteesin kanssa, vahvistusharhan voi välttää esimerkiksi käyttämällä koko korpusta ja kaikkia siitä nousevia relevantteja löytöjä hypoteesin testaamiseen. Mikäli koko korpusta on mahdoton käyttää, tutkijan pitäisi kuitenkin pyrkiä välttämään aineiston tietoisista valikointia. Ongelman ratkaisemiseksi on mahdollista käyttää esimerkiksi satunnaista otantaa (mp.), tai systemaattista satunnaisotantaa, kuten tässä tutkimuksessa on tehty vahvistusharhan välttämiseksi. Toki täytyy myös muistaa, että korpus itsessään edustaa aineiston valintaa ja karsimista (McEnery & Hardie 2014: 15).

Käytännössä kumpaakin korpusaineistoa on hyödynnetty konkordanssityökalulla korpuspohjaisen tutkimusmenetelmän periaatteiden mukaisesti niin, että kummastakin aineistosta haettiin SAADA-verbin esiintymiä. Konkordanssityökalu mahdollistaa minkä tahansa mittaisen merkkijonon hakemisen korpuksesta: sanan, osan sanasta tai fraasin. Lisäksi annotoidusta korpuksesta voi KWIC-työkalulla hakea sinne annotoitua kielellistä tietoa, kuten tässä tutkimuksessa tehtiin. Hakutulos näytetään konkordanssina yksi hakutulos per rivi niin, että hakutuloksen ympärillä näkyy kontekstia ennen ja jälkeen. Tällaista menettelyä kutsutaan nimellä *key word in context* (KWIC) (vapaasti suomennettuna *avainsana kontekstissa*). (McEnery ja Hardie 2014: 35.) Tässä tutkimuksessa käytetyt aineistot on kerätty KWIC-työkalulla niin, että kontekstin laajuus on yksi virke.

Oppijansuomen aineiston keräyksessä ICLFI-korpuksesta käytettiin Korp-käyttöliittymän versiota 5.0.10 (Borin, Forsberg & Roxendal 2012). Konkordanssi järjestettiin oletuksen sijaan satunnaisesti. Satunnaistaminen takaa sen, että systemaattisella satunnaisotannalla kerätty aineisto on mahdollisimman sattumanvarainen eikä vahvistusharhaa pääse syntymään. Tämän jälkeen aineistoksi poimittiin ensimmäiset 513 SAADA-verbin esiintymää virkekonteksteineen. Tarkka hakulauseke Korpin edistyneessä haussa on seuraava¹⁸:

```
[lemma = "saada" %c & ( _text_cefrlevel = "B2" | _text_cefrlevel = "C1" | _text_cefrlevel = "C2")]
```

Lauseke hakee aineistosta SAADA-verbin lemmana (“lemma = saada”), eli kaikki sen esiintymät kaikissa muodoissaan. Hakulausekkeessa isot ja pienet kirjaimet on samaistettu, joten kirjainkoolla ei ole väliä (%c). Loppulauseke “(_text_cefrlevel = "B2" | _text_cefrlevel =

¹⁸ Vaikka haku on tehty satunnaistamalla, täsmälleen samaan hakutulokseen pääsee käyttämällä random seedia 6453606.

"C1" | _.text_cefrlevel = "C2")" määrittää lisähakuehdoksi taitotason B2 tai C1 tai C2, jolloin konkordanssissa näkyy kaikkien haettujen taitotasojen SAADA-verbien esiintymät.

Käännösuomen korpus on annotoimatonta raakatekstiä, mutta tutkimuksessa käytetään Jantusen (2004) *Connexorin Machineese Phrase Tagger* -ohjelmalla annotoimaa versiota. Tekstistä etsittiin AntConc-ohjelman versiolla Windows 64-bit 3.5.7 (Anthony 2018) kaikki SAADA-verbin lemmat, ja systemaattisen satunnaisotannan avulla aineistoksi kerättiin joka 30. hakuosuman aloittaen 1. konkordanssirivistä. Mikäli otantamenetelmällä osuma sattui *saada*-sanamuotoon, valittiin seuraava (eli seuraavan 31:n 30:n sijaan) osuman, joka on kyseisen sanamuodon lemma-taggi. Tällä tavoin on siis kerätty joka 30. lemma, ei joka 30. osumaa. Näin saatiin 513 SAADA-verbin esiintymää virkekonteksteineen. Tämän jälkeen jouduttiin kasvattamaan aineistoa. 300 virkettä kerättiin lisää niin, että aloitin hakutuloksesta 15 ja tästä eteenpäin kerättiin joka 30. lemma, kunnes uudet 300 virkettä oli kasassa. Näin ollen nättiivisuomen aineiston koko on 813 virkettä. Aineisto on aineistonkeruuhetkellä järjestetty genrejen mukaisesti, ja aineistontäydennyksestä johtuen tulokset saattavat painottua hieman alkupään teksteihin.

5.3 Intuition hyödyntäminen korpustutkimuksessa

Vaikka korpustutkimus pohjautuu vahvasti aineistoon, siinä on läsnä myös intuitiivisia piirteitä. Aineistopohjaisuutta ja intuition perustuvaa menetelmää ei siis tulisi nähdä vastakkaisina, vaan jopa toisiaan täydentävinä tutkimusmenetelminä (ks. Jantunen 2009a). Vaikka tutkimuksen keskiössä on autenttinen kieli, ei tutkimuksen tekeminen onnistuisi ilman tutkijan tekemiä intuitiivisia oletuksia: esimerkiksi tutkimusaiheen ja aineistojen valinta perustuu tutkijan intuition kielestä (Stubbs 2002: 9; Jantunen 2009a: 107; Pirkola 2016: 39).

Myös aineiston käsittely edellyttää tutkijan intuitiota (Jantunen 2009a: 107). Jantunen (2004: 211) ja Pirkolan (2016: 39) tapaan tässäkin tutkimuksessa nojataan tutkijan intuition etenkin, kun SAADA-verbin esiintymiä ryhmitellessä sen eri merkitysten mukaan. Tässä apuna on Kielitoimiston sanakirja, mutta loppujen lopuksi luokittelu tapahtuu tutkijan intuition ja tilikinnan varassa. Lisäksi intuitiota hyödynnetään hyvin vahvasti semanttisen primingin analyysissä. Sanojen luokittelu perustuu siihen, mitkä sanat intuitiivisesti kuuluvat yhteen – ja toisaalta siihen, kuinka intuitiivisesti tutkija ymmärtää sanan merkityksen (ks. mp.).

Intuitiivista otetta tutkimuksen laadullisissa osioissa ei ole siis tarpeen nähdä tutkimuksen heikkoutena, vaan pikemminkin sen ominaispiirteenä (Pirkola 2016: 39). Kaikkeen sanan

merkitykseen liittyvään tietoon (myös intuitiiviseen) tulisikin suhtautua lähtökohtaisesti kiinnostuneena (Vanhatalo 2003: 362). Tässä tutkimuksessa yhdytään Jantusen (2009a) ja Pirkolan (2016: 40) näkemyksiin siitä, että vahvaa aineistoon sekä intuitioon nojautumista ei tulisi nähdä toisilleen vastakkaisina, vaan myös aineistopohjaisessa korpustutkimuksessa tutkijan intuitio kielestä on arvokasta.

5.4 Tilastolliset menetelmät

Tässä tutkimuksessa tarkasteltavat muuttujat (morfologinen ja semanttinen priming) ovat luokittelu- eli nominaaliasteikollisia. Kaikista mahdollisista muuttujien tyypeistä luokitteluas- teikon muuttujat ovat vähiten informatiivisia (Valli 2001: 32), mutta niillekin on mahdollista laskea erilaisia arvoja: tässä tutkimuksessa tilastollisina menetelminä hyödynnetään z-testiä ja X^2 -riippumattomuustestiä¹⁹.

Toinen tässä tutkimuksessa käytetty tilastollinen testi on kahden riippumattoman otoksen z-testi. Z-testillä testataan suhteellisten arvojen erojen tilastollista merkitsevyyttä: onko ero tilastollisesti melkein merkitsevä, merkitsevä vai erittäin merkitsevä. Merkitsevyyden aste kertoo, kuinka suurella todennäköisyydellä tulos johtuu sattumasta. Z-arvon ollessa vähintään 1,96 tulos on melkein merkitsevä, ja tällöin on 5 % ($p < 0,05$) mahdollisuus, että tulos johtuu sattumasta. Kun z-arvo on vähintään 2,58, tulos on merkitsevä, ja tällöin sattuman mahdollisuus on 1 % ($p < 0,01$). Kriittinen z-arvo erittäin merkitsevälle tulokselle on 3,29, jolloin on 0,1 % ($p < 0,001$) todennäköisyys, että tulos johtuu sattumasta. (Kanji 1993: 161.) Tässä tutkimuksessa z-testausta käytetään tutkimuskysymysten yksi, kaksi ja kolme osalta testaamaan SAADA-verbin merkitysluokkien, morfologisen primingin ja semanttisen primingin suhteellisten osuuksien eroja oppijan- ja natiivisuomen aineistoissa. Testi näin ollen nostaa esiin edellä mainittujen muuttujien merkittävimmät erot aineistojen välillä.

X^2 -riippumattomuustestiä varten täytyy ensin tehdä ristiintaulukointi, jolla selvitetään, kuinka kaksi muuttujaa vaikuttavat toisiinsa. Ristiintaulukoinnissa muuttujat esitetään samassa taulukossa niin, että toinen asetetaan riveille (tässä tutkimuksessa semanttinen priming) ja toinen sarakkeille (tässä tutkimuksessa semanttinen priming). (Heikkilä 2005: 210.) Jo ristiintaulukoinnista voi päätellä, onko muuttujien välillä yhteyttä, mutta tilastollisen yhteyden todentamista varten tarvitaan X^2 -riippumattomuustesti. Tässä testissä nollahypoteesi on, että muuttujien välillä ei ole riippuvuutta, ja testi laskee nollahypoteesin mukaiset odotetut frekvenssit

¹⁹ Testistä käytetään myös nimitystä khiin neliö -testi (Heikkilä 2005: 212).

(*expected counts*). Odotetut frekvenssit ovat arvoja, joiden mukaan muuttujien välillä ei olisi lainkaan riippuvuutta. Tämän jälkeen testi laskee testisuureet, joiden arvo on sitä suurempi, mitä enemmän odotettu arvo poikkeaa havaitusta arvosta. Testisuureen ja vapausasteen (rivien lukumäärä - 1 x sarakkeiden lukumäärä - 1) avulla testi laskee merkitsevyytason, joka kertoo, kuinka suurella todennäköisyydellä muuttujien välillä ei ole riippuvuutta. (Heikkilä 2005: 212.) Ristiintaulukointia ja X^2 -riippumattomuustestiä käytetään vastaamaan tutkimuskysymykseen neljä, jossa tarkastellaan muodon ja merkityksen välistä suhdetta.

6 ANALYYSI

Ensimmäisessä alaluvussa eritellään SAADA-verbin merkitysluokittelun perusteita ja tuloksia. Luokittelussa on käytetty apuna Kielitoimiston sanakirjaa (KS s.v. *saada*), jonka mukaan SAADA-verbillä on 15 eri merkitystä (ks. liite 1), joista tämän tutkimuksen aineistoissa esiintyy 14. Tämän jälkeen siirrytään tutkimuskysymysten mukaisesti ensin morfologisen primingin ja sitten semanttisen primingin analyysiin. Lopuksi tarkastelen morfologisen ja semanttisen primingin välistä korrelaatiota.

6.1 SAADA-verbin polysemia

6.1.1 Merkitysluokittelun periaatteet

Vaikka luokittelu on tehty valmiisiin Kielitoimiston (KS s. v. *saada*) määrittelemiin luokkiin, moni tapaus on silti ollut hyvin tulkinnanvarainen ja päätös luokasta on tehty pienten nyanssien varassa. Luokittelussa tutkijan intuitio on siis ollut merkittävässä asemassa. Joku toinen voisi tehdä luokittelun hieman eri tavalla, mutta luokittelussa on pyritty systemaattisuuteen ja johdonmukaisuuteen. Mikäli merkitysluokka ei ole ilmennyt suoraan virkkeen laajuisesta kontekstista, on tarkistettu virkettä laajempi konteksti.

Yleisiä linjauksia ja eri merkitysluokkien selkeitä rajoja on vaikeaa määritellä, koska kaikki virkkeet on luokiteltu yksittäistapauksina, jokainen omassa kontekstissaan firtiläisen periaatteen mukaisesti. Luokittelun perusteille on kuitenkin johdonmukaiset suuntaviivat, joita on pyritty noudattamaan. Koska oppijansuomen aineiston koko on 513 virkettä ja natiivisuomen aineiston koko 813 virkettä, ei kaikkia virkkeitä pysty esittämään. Tämän vuoksi esimerkeissä pyritään kuvastamaan kustakin luokasta peruserkitys, mahdolliset erikoismerkitykset, sekä merkittäviä rajatapauksia, jotka ovat vaikuttaneet luokitteluun. Esimerkit on pyritty valitsemaan niin, että ne edustaisivat aineistoja mahdollisimman kattavasti.

Merkitysluokkien ja niiden periaatteiden esittelyn jälkeisessä alaluvussa esitetään, kuinka aineistot jakautuvat luokkiin. Koska luokkia on paljon, kuvioiden hahmottaminen on helpompaa luokkien läpikäymisen jälkeen. Seuraavassa alaluvussa on hahmottamisen helpottamiseksi merkitysluokan numeron viereen on annettu luokan merkitys tiivistettynä.

Merkitysluokka 1: saada jotain omistukseen tai käyttöön

Ensimmäisessä luokassa jotakin siirtyy jonkin tai jonkun omistukseen tai käyttöön. Kyse voi olla esineestä, hyödykkeestä, oikeudesta tai tulkintani mukaan myös palvelusta (KS s.v. *saada*):

1. tapahtumasta jossa jk esine, hyödyke, oikeus tms. siirtyy jkn t. jnk omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jk tulee. *Sai isältään kellon ja rahaa. Saada lahja, perintö. Ei saanut pankista lainaa. Saada hyvä virka, työpaikka. Saada jtk omakseen, käyttöönsä, lainaksi. Saada jtk perinnöksi (t. perintönä), lahjaksi (t. lahjana), palkinnoksi (t. palkintona). Ei saanut koko päivänä syödäkseen. Sai vastaanottilijalta tulen tupakkaansa. Ovatko kaikki saaneet kahvia? Saisinko [= antaisit(te)ko minulle] lasin vettä.*

Erik.

a. jnk lähetetyn vastaanottamisesta. *Saada kirje, viesti jkltä. Saada vastaus tiedusteluunsa. Kuitata lähetys saaduksi.*

b. jnk oikeuden, mahdollisuuden, etuisuuden tms. tulemisesta jkn osaksi. *Saada lupa, oikeus jhk. Sai kunnian toimia puheenjohtajana. Saada virkavapautta, lomaa. Sai eron vaimostaan. Pyysi ja sai puheenvuoron. Saat vielä yhden tilaisuuden yrittää. Saada hyvää palkkaa.*

Kuten jo Kielitoimiston sanakirjan oman määritelmän ja esimerkkien mukaan voi havaita, merkitysluokka 1 on hyvin monipuolinen – se pitää sisällään niin konkreettisten asioiden saamista haltuun tai käyttöön kuin myös abstraktien seikkojen saamisen, kuten lainan ja viran. Oppijansuomen aineistoesimerkit 1 ja 2 ja natiivisuomen aineistoesimerkit 3 ja 4 ovat selviä luokkaan 1 kuuluvia tapauksia:

[1] Jokainen ihminen **saa** myös suklaamunat. (O)²⁰

[2] Nykyään Venäjällä on vaikea **saada** hyvän työn. (O)

[3] Pienempi **sai** huuliharpun, joka parkaisi ilkeästi. (N)

[4] Isä oli nuorena voimistellut ja urheillut, **saanut** ihan palkintojakin, ja äiti oli lapsena voittanut koulunsa juoksukilpailut. (N)

Ongelmallista merkitysluokassa 1 on rahan tms. saamiseen liittyvät tapaukset, koska osin ne voisi tulkita myös luokkaan 4. Tapaukset onkin määritelty tapaukset niin, että luokassa 4 korostuu se, että saaminen on onnistumista jostain. Merkityksestä täytyy siis tulla selvästi ilmi, että onnistumista ennen on täytynyt nimenomaan aktiivisesti yrittää tavoitella jotain (kuten rahaa) haltuunsa. Tähän palataan tarkemmin luokan 4 käsittelyssä. Luokassa 1 sen sijaan rahaa saa olematta juuri itse aktiivinen. Tällaista tilannetta kuvaa oppijankielen esimerkki 5. Lisäksi luokkaan 1 menee rahan tavoittelu ilman sen saamista – tällöinhän kyse ei ole onnistumisesta. Tällaista tapausta kuvaa natiivisuomen aineistoesimerkki 6.

[5] Minun tyttäreni ja hänen miehensä **saisivat** rahaa minulta. (O)

[6] Mutta Askto yritti **saada** rahaa vanhemmilleen, joilla oli vaikeuksia. (N)

²⁰ Esimerkkien perässä kirjain O merkitsee oppijansuomen esimerkkiä ja N natiivisuomen esimerkkiä.

Sen lisäksi, että luokkien 1 ja 4 välillä on niiden kuvausten vuoksi tasapainottelua, myös valmiiksi annetut esimerkit johtavat välillä leksikaalisella tasolla harhaan, kuten oppijankielen aineistosta poimitussa esimerkissä 7. Tässä on hyvä esimerkki siitä, miksi luokittelussa on tärkeää tarkastella verbiä sanatasoa syvemmin. Luokan 4 merkityksessä eräs esimerkki on *saada päävoitto arpajaisissa*, mikä on semanttisesti hyvin samankaltainen aineistoesimerkki 7:n kanssa. Luokan 4 kuvausesimerkissä saadaan kuitenkin päävoitto, eli onnistutaan hankkimaan itselleen jotain (mitä toisaalta voi kyseenalaistaa, onko kyse varsinaisesti onnistumisesta), ja aineistoesimerkissä 7 saadaan arpalippuja – eli esineitä tai tulkintatavasta riippuen hyödykkeitä, eikä siis onnistuttu saamaan voittoa. Myös oppijansuomen aineistoesimerkki 8 on oiva havainnollistamaan, kuinka tarkkaa polyseeminen jaottelu voi olla: esimerkissä 8 kyse on palvelun tai palvelua tarjoavan henkilön saamisesta jonkun käyttöön, jolloin kyse on luokasta 1. Mikäli kyse olisi ollut suoraan hoidon saamisesta, olisi se toimenpiteenä mennyt luokkaan 2.

[7] Tänä **sain** viisi arpalippua, mutta ilman voittoa. (O)

[8] Lapseni sairastui yöllä korvatulehdukseen ja minulla ei ole mahdollista **saada** hänelle hotajaa huomiseksi. (O)

Merkitysluokkien 1 ja 4 väliseen rajanvetoon liittyvät myös tapaukset, joissa saadaan palkkaa. Palkanmaksun voi nähdä joko etuutena (luokka 1b) tai toimeentulon hankkimisena (luokka 4). KS on kuitenkin määritellyt luokan 1 esimerkissä, että 'saada hyvää palkkaa' on etuus. Pohdittavan arvoista onkin, muuttuisiko merkitys etuudesta toimeentuloksi ilman määritettyä *hyvä*. Luokan 4 esimerkeissä ei kuitenkaan ole viitteitä palkan saantiin, joten palkan saanti on luokiteltu merkitykseen 1 (oppijansuomen aineistoesimerkki 9 ja natiivisuomen aineistoesimerkki 10).

[9] Kun opiskelee, oppii asioita ja **saa** parempaa palkkaa. (O)

[10] Lohtunani oli, että nyt **sain** rahapalkkaa. (N)

Merkitysuokalla 1 on erikoismerkitykset a 'jonkin lähetetyn vastaanottamisesta' ja b 'jnk oikeuden, mahdollisuuden, etuisuuden tms. tulemisesta jkn osaksi'. Näistä merkitys 1a (oppijansuomen aineistoesimerkit 11 ja 12 ja natiivisuomen aineistoesimerkit 13 ja 14) on melko yksiselitteinen, eikä aineistoissa ollut tähän liittyviä ongelmatapauksia. Sen sijaan merkitys 1b on paikoin ongelmallinen, sillä se on paikoittain päällekkäinen luokkien 9 ja 11 kanssa. Selkeitä tapauksia luokan 1b-kohdalla edustavat oppijansuomen aineistoesimerkki 15 ja natiivisuomen aineistoesimerkki 16.

- [11] Olin oikein iloinen, kun olen **saanut** kirjeen sinulta! (O)
 [12] Minähän **sain** aika kauan sitten tuo sähköposti sinulta, missä olit kirjoittanut siitä kurssista. (O)
 [13] Jack kosii Armia uudelleen ja **saa** myöntävän vastauksen. (N)
 [14] Seuraavana päivänä Esku **sai** postia. (N)
 [15] Kirjojen kopioiminen maailma ympäri on hyödyllinen myös siksi, että enemmän ihmisiä **saa** pääsyä harvinaisiin teoksiin. (O)
 [16] Tero taisi nyt **saada** lomaa remontista. (N)

Merkitysluokka 2: saada jokin käsittely tai toimenpide

Merkitysluokassa 2 KS (s. v. *saada*) kuvaa SAADA-verbiä seuraavasti:

2. tulla, joutua jnk toimenpiteen, käsittelyn, kohtelun tms. kohteeksi. *Saada syyte, haaste oikeuteen. Saada apua, hoitoa. Saada valmennusta, opetusta. Hyvän kasvatuksen saanut lapsi. Ehdotus sai innostuneen vastaanoton.*

Erik.

a. joutua kurituksen, pahoinpitelyn, moitteiden tms. kohteeksi. *Saada selkäsauna, selkäänsä. Saada keppiä t. kepeitä, piiskaa t. piiskasta. Pojat saivat [kuritusta] oikein isän kädestä. Sai tytöltä ympäri korvia. Saada jkltä isku, potku. Saada nuhteita, moitteita, sapiskaa, kyytiä. Sai mitä ansaitsikin. Saada nenälleen t. nenilleen tulla nolatuksi.*

Ark. Saada turpiinsa, kuonoonsa, köniinsä. Siitäs sait, kun et uskonut. Saada siipeensä.

b. tulla tuomituksi rangaistukseen. *Saada rangaistus, tuomio rikoksesta. Saada vankeutta, arestia, sakkoa. Sai neljä vuotta (vankeutta) petoksesta.*

c. tulla, joutua kiittävän, moittivan tms. suhtautumisen kohteeksi. *Saada osakseen kiitosta, myötätuntoa, halveksuntaa, pilkkaa. Mainetta saanut taiteilija.*

d. (koulu)arvosteluun t. arviointiin liittyen. *Saada hyvä todistus. Sai kokeessa nelosen. Saada ehdot. Mäkihyppääjä sai kaikilta tuomareilta 18 pistettä. Kirja sai hyvät arvostelut.*

Esimerkkien lisäksi tähän luokkaan tulkitaan tarjousten (oppijansuomena aineistoesimerkki 17, natiivisuomen aineistoesimerkki 18), potkujen (oppijansuomen aineistoesimerkki 19), siirron (natiivisuomen aineistoesimerkki 20), (taloudellisen) tuen (oppijansuomen aineistoesimerkki 21), palautteen (oppijansuomen aineistoesimerkki 22) ja huomion saamisen (oppijansuomen aineistoesimerkki 23). Muutoin merkitysluokka 2 on selkeä, eikä siinä ole päällekkäisyyksiä muiden luokkien kanssa.

- [17] 1972 minä **sain** tarjouksen työpaikasta Helsingissä, pienessä galleriassa. (O)
 [18] [- -] Koska muita tarjouksia ei ole **saatu** ja hinta ylittää asetetun lähtöhinnan, on kaupan ollut irtaimisto myyty kokonaisuutena Santo Oy:lle. (N)
 [19] Syyllinen **saa** potkut liitoltamme [- -]. (O)
 [20] Maaliskuussa Tiina oli **saanut** siirron työpaikkansa Lontoon konttoriin. (N)
 [21] Jos tarvitsen tukea, voin **saada** sitä vanhemmiltani. (O)
 [22] Uutiset Talous- ja Ulkomaat-osastoista saivat paljon kommentteja ja herättivät rajua keskustelua, muutamat artikkelit **saivat** noin 30 kommenttia ja noin 160 ihmistä suosittelee artikkelia ystävilleen. (O)
 [23] Total Balalaika Show' oli kuvattu yhteistyössä Aleksandrovinnimisen yhtyen kanssa, joka oli esitetty Suomessa ja myös Kannien elokuvafestivaaleissa ja **saanut** suurta huomiota. (O)

Merkitysluokka 3: saada jotain jonkin osaksi

Luokka 3 määritellään seuraavasti:

3. ilmaantua t. tulla jnk osaksi t. hyväksi. *Kuu saa valonsa auringosta. Saada väriä ihoonsa. Lapsi saa hampaista. Ravinnosta saatavat hivenaineet. Saada rohkeutta, voimaa jstak. Sai elämältä paljon. Saada uskonnollinen herätys. Sai päähänsä hullun idean. Saada harmia jstak. Sai surmansa kolarissa. Saatiin hyvät ilmat. Huhu sai siivet. Tapaus sai luonnollisen selityksen selittyi luonnolliseksi. Saada alkunsa alkaa; aiheutua, syntyä.*

Erik.

a. jälkeläisen syntymästä. *Nainen sai lapsen, pojan. Mies sai perillisen, lapsenlapsen. Sai lapsen tyttöystävänsä kanssa. Koira saa pentuja.*

b. sairauteen, oireeseen t. vammaan liittyen. *Saada tauti, kuumetta, päänsärkyä. Sai ruoasta mahanpuruja. Saada sydänkohtaus, halvaus. Sai lasinsirusta haavan.*

c. ark. hämmästyttä, raivostumista, pelästymistä, inhoamista tms. merkitsevissä ilmauksissa. *Saada kohtaus, halvaus, sätky, hepuli ks. kohtaus jne. Saan näppylöitä, kun näenkin hänet.*

d. muuttua, kehittyä jonkinlaiseksi. *Valukappale saa muotonsa muotissa. Uuden merkityksen saanut sana. Asia sai yllättävän käänteen.*

Kolmannen merkitysluokan tapaukset eivät ole aina täysin selviä, ja luokan perusmerkityskin jakautuu useaan alamerkitykseen. Tällaisia alamerkityksiä ovat esimerkiksi 'saada kokemuksia' (oppijansuomen aineistoesimerkki 24 ja natiivisuomen aineistoesimerkki 25), 'saada itsevarmuutta' (oppijansuomen aineistoesimerkki 26), 'saada motivaatiota' (oppijansuomen aineistoesimerkki 27), 'saada henkistä valoa tai lohdutusta' (oppijansuomen aineistoesimerkki 28), 'saada (henkistä) tukea' (oppijansuomen aineistoesimerkki 29 ja natiivisuomen aineistoesimerkki 30), 'saada ikävä'* (oppijansuomen aineistoesimerkki 31) ja 'saada (elämältä) paljon' (oppijansuomen aineistoesimerkki 32):

[24] Luotan että **saan** usea uusia kokemuksia ja opin paljon lääkäriksi olemisesta. (O)

[25] Etelä-Pohjanmaan keskussairaalan malli on toistaiseksi ainutlaatuinen ja myös siitä **saadut** kokemukset ovat hyvät. (N)

[26] Huomasin myös, että olin **saanut** kirjoittamiseen lisää itsevarmuutta. (O)

[27] Mutta yritän keskittyä siihen kunnolla, **sain** taas motivaatiota, en haluata tuhjata sitä. (O)

[28] Syrjäinen ja synkkä Suomussalmi **saa** teatterista henkistä valoa ja lohdutusta. (O)

[29] **Saanut** tukea rakkaudesta, Kai päättää tappaa demonin. (O)

[30] Kustaa Vaasa vastaan Hansa Kustaa Vaasa ei **saanut** Lyypekin tukea unioninvastaiselle kapinalleen ilmaiseksi. (N)

[31] Voi olla, että se oli pääperuste, miksi hänelle **sai** ikävä. (O)

[32] Vaikka minulla ei ole poikaystävä tänä vuonna, elämä on kaunista ja olen **saanut** paljon. (O)

Luokan 3 erikoismerkityksistä aineistoissa esiintyy a. 'jälkeläisen syntymä' (oppijansuomen esimerkki 33 ja natiivisuomen esimerkki 34), b. 'sairauteen, oireeseen t. vammaan liittyen' (oppijansuomen esimerkki 35 ja natiivisuomen esimerkki 36) ja d. 'muuttua, kehittyä jonkinlaiseksi' (oppijansuomen esimerkki 37 ja natiivisuomen esimerkki 38). Erikoismerkitystä c.

'hämmästyttä, raivostumista, pelästymistä, inhoamista tms. merkitsevissä ilmauksissa' ei ollut kummassakaan aineistossa.

- [33] Valtion täytyy tehdä lapsien saaminen tuottava, niin että nuoret aikuiset haluavat **saada** vauvoja. (O)
 [34] Mutta kenties piankin tulee elämäni nuori nainen, jonka kanssa **saamme** lapsia. (N)
 [35] Kyläläiset pesivät silloin Bedrichin vamman vedessä, joka oli likaista ja sitä varten hän **sai** kasvoonsa tulehduksen, joka harmitti häntä koko elämän ja varmaankin aiheutti hänen kuolemansa. (O)
 [36] Ehkä olin **saamassa** vatsataudin. (N)
 [37] Hän riemuissaan asioiden **saamasta** käänteestä, mutta kanssaihmisillä on vaikeuksia sopeautua kuolevan miehen kuolematomuuteen. (O)
 [38] Toisin työttömyys minäkuvanakin voi **saada** useita erilaisia muotoja. (N)

Merkitysluokka 4: onnistua saamaan jotain

Luokalla 4 on myös perusmerkityksensä lisäksi useita merkityksiä:

4. onnistua hankkimaan jtk itselleen, käyttöönsä, yhteyteensä tms., tavoittaa, saavuttaa. *Sai kirjan kymppillä, alennuksella. Sai halvalla hyvän auton. Odota, että ehdin saada jotakin ylleni. Luki kaiken, minkä käsiinsä sai. Saada pakeneva kiinni. Sai kiinni köydestä. Saada jku valtaansa. Saada ote vastustajasta. Saada kosketus, yhteys, tuntuma jhk. Saada osuma maalitauluun. Juoksija sai hyvän lähdön onnistui lähdössä hyvin. Sai kultamitalin viimeisellä heitolla. Saada päävoitto arpajaisissa.*

Erik.

- a. onnistua elatuksen, sadon, saaliin tms. hankkimisessa. *Saada toimeentulonsa kaupasta. Pellosta saatiin hyvä sato. Metsästäjä sai hyvän saaliin. Onki tunnin saamatta mitään.*
- b. onnistua hankkimaan jku seurakseen t. ystävyys- tm. suhteeseen. *Saada jku toverikseen, saattajakseen. Hän on saanut paljon ystäviä, vihamiehiä kirjoituksillaan.*
- c. onnistua hankkimaan jku puolisoon. *He saivat toisensa. Arka mies ei saa koreata akkaa SL.*
- d. ark. päästä (naisen kanssa) yhdyntään. *Saitko sä siltä? Saada naista, miestä.*
- e. ark. orgasmin saavuttamisesta. *Saada yhdynnässä. Saitko?*

Kuten merkitysluokan 1 käsittelyssä tuli ilmi, toisinaan rajanveto merkitysluokkien 1 ja 4 välillä on haasteellista. Alla on kuitenkin esimerkkejä (oppijansuomen aineistoesimerkit 39–42 ja natiivisuomen aineistoesimerkit 43–47), joissa kyse on selkeästi luokasta 4:

- [39] Tulen lentämällä, tiedäthän että **saan** lipun halvalla kun siskoni tekee työtä lentokentällä. (O)
 [40] Muistan, että siskoni ja minä olimme vähän matkaa tien päällä edelläpäin ja huusimme sille työlle, että hän juoksisi kovempaa, muuten pojat **saavat** meidät kiinni. (O)
 [41] Sellaiset pirteet jo lapsena autoi hänelle **saada** enemmistöä yliotteen. (O)
 [42] Hänen esivanhemmat **saivat** toimeentulonsa saksien alalta. (O)
 [43] Kellarin luukku on pihalle auki ja trumpetti kuuluu korviin niillä rajoilla, etten **saa** melodiasta kiinni, klassista standardia. (N)
 [44] Joku hänen toverinsa on muka **saanut** käsiinsä tehtaan, mutta ei ole kunnan koneita. (N)
 [45] [– –] sen minä ymmärrän ja hyväksyn, että isä **sai** sinun neitsyytesi. (N)
 [46] [– –] suurtuottajat **saivat** ensimmäisen voittonsa jo neljännesvuosisata ennen kotipolito-oikeuden täydellistä lakkautamista. (N)
 [47] Teistähän kerrotaan, että **saatte** yhteyden jopa Troijan Helenaan. (N)

Epäselviä tapauksia, jotka on luokiteltu merkitysluokkaan 4, ovat muun muassa tutkinnon ja ajokortin saaminen (oppijansuomen esimerkit 48 ja 49). Näiden tapausten voisi ajatella

olevan asioita, jotka siirtyvät jonkun hyödynnettäväksi, mutta tällaisten asioiden saamisen tulkitaan kuitenkin vaativan aktiivista tavoittelua ja onnistumista, mikä on luokkien 1 ja 4 merkittävimpanä eroavaisuutena.

[48] Lopuksi **sain** logopedin tutkinto vuonna 2017. (O)

[49] Tiger ei ole aivan tavallinen ihminen, hän ei voi olla työssä, **saada** ajokorttia, mennä armeijaan. (O)

Luokka 4 risteää paikoittain myös luokan 8 kanssa, sillä molemmissa kuvauksissa mainitaan 'onnistuminen'. Luokassa 4 onnistuminen koskee sitä, että onnistuu hankkimaan jotain itselleen, luokassa 8 sitä, että onnistuu jonkin asian tekemisessä. Natiivisuomen aineistoiesimerkki 50 kuvastaa tällaista tilannetta. Esimerkki on luokiteltu merkitykseen 4, koska onnistumisen tulkitaan kohdistuvan siihen, että jotain saadaan itselleen tai omaan käyttöön.

[50] Miehet eivät irvistelleet, mutta kun **sain** pullon suulleni, pontikan maku pisti kurkkuni nykimään. (N)

Aineistoissa esiintyy myös luokan 4 erikoismerkityksiä: a. 'onnistua elatuksen, sadon, saaliin tms. hankkimisessa' (natiivisuomen aineistoiesimerkit 51 ja 52), b. 'onnistua hankkimaan joku seurakseen tai ystävyys- tm. suhteeseen' (oppijansuomen aineistoiesimerkit 53 ja 54) c. 'onnistua hankkimaan joku puolisoon' (natiivisuomen aineistoiesimerkki 55), d. 'päästä (naisen kanssa) yhdyntään' (natiivisuomen aineistoiesimerkki 56). Merkitystä e. 'orgasmin saavuttamisesta' ei ollut kummassakaan aineistossa.

[51] Hetken päästä hän **sai** pienen särjen ja suunnitteli pyydystävänsä seuraavaksi kalan puukolla ja makasi laiturilla mahallaan puukko kädessä, mutta kalaa ei sillä menetelmällä tullut. (N)

[52] Se aleni alle puoleen entisestään 36 pennistä kilolta 17 penniin, ja kun kahvin tuontitulli Suomessa pysyi 40 penninä kilolta, nousi salatuonnista Suomessa **saatava** voitto huomattavasti. (N)

[53] Mutta kohta hän tutustui toisiin opiskelijoihin ja hän **sai** monta ystävää. (O)

[54] Jos oli rikas, **sai** ystäviä... (O)

[55] Gil oli rukoillut koko elämänsä voimalla, että jokin ihme tapahtuisi ja hän **saisi** tämän vaalean naisen ikiajoiksi omakseen. (N)

[56] - Sitten siltä voi vaikka **saada!** (N)

Merkitysluokka 5: saada tietää tai ymmärtää

Luokalla 5 on vain sen perusmerkitys: 'jonkin tulemisesta jonkun aistimusten, tiedon tai ymmärryksen piiriin':

Sai pakenijan näkyviinsä. Saada korviinsa, tietoonsa ilkeitä huhuja. Saada tieto päätöksestä. Sai vihiä suunnitelmasta. Saada käsitys, aavistus, vaikutelma, selko jstak. Ei saanut selvää käsialasta.

Mrekitysluokka on selkeä, melko rajatulla merkityksellä (vrt. luokka 1). Aineistosta nousevat tapaukset ovat samankaltaisia kuin KS:n esimerkeissä (oppijansuomen esimerkit 57 ja 58 sekä natiivisuomen esimerkit 59 ja 60), mutta tulkinnanvaraisia tapauksia löytyy tästäkin luokasta (oppijansuomen esimerkki 61).

- [57] Kulttuurikurssilla olen **saanut** paljon tietoja Suomesta. (O)
 [58] Olen **saanut** vaikutuksen, että nykyinen maailman tilanne näyttää tavalliseen verraten päinvastaiselta. (O)
 [59] On tärkeää **saada** käsitys myös lapsen älyllisestä kehityksestä. (N)
 [60] Se **sai** selville, että satiainen on millin mittainen otus. (N)
 [61] En tarkoita että ymmärtäisin heti kirjoitettua mutta sitä **saa** kelata monta kertaa uudestaan. (O)

Esimerkki 61 voisi olla myös luokassa 12 ('joutua tekemään jotain'), mutta fraseologisten periaatteiden mukaisesti virke on analysoitu kokonaisuudessaan. Tässä tapauksessa *saa kelata monta kertaa uudestaan* tarkoittaa asian pohtimista useaan kertaan ennen ymmärtämistä, eli kyse on 'ymmärtämistä' tarkoittavasta metaforasta.

Harkitsemista vaativat myös tapaukset 'saada tietää', jolloin kyse voisi olla myös 'mahdollisuudesta'. Tällaisissa tapauksissa on päädytty siihen, että niissä painottuu tiedon saanti (asian tuleminen ymmärryksen piiriin) eikä niinkään mahdollisuus tietämiseen (oppijansuomen aineistoesimerkki 62 ja natiivisuomen aineistoesimerkki 63).

- [62] Hän ei koskaan antaisi Timon olla jos hän **saisi** tietää että Timo pitää Kaisasta. (O)
 [63] Minun pitäisi **saada** tietää paljonko hän painaa. (N)

Luokkaan 5 on päätyneet myös mm. seuraavanlaiset 'ymmärtämistä' kuvaavat tapaukset (oppijansuomen aineistoesimerkit 64–67 ja natiivisuomen aineistoesimerkit 68–70):

- [64] Myös, jos minä en ymmärrä joitakuita sanoja, se on vaikea minusta **saada** langan päästä kiinni. (O)
 [65] Minä esimerkiksi alussa pyrin **saaman** yleiskuvan vierasta kielestä, sen rakenteista, syntäksistä, kieliopin pääsäännöistä ja vasta sitten aloitan opetella sanoja, ilmaisuja. (O)
 [66] [– –] **sain** vielä tukea entiselle käsitykselle, että [– –] kieliopetuksessa voi onnistua erittäin hyvin. (O)
 [67] Minulla on pari todella hyvää keittokirjaa, sieltä **saa** varmasti hyviä ajatuksia [– –]. (O)
 [68] Eräänä päivänä tiikeri **sai** lehmän hajun ja lähti ajamaan sitä takaa. (N)
 [69] En **saa** niistä tolkkua. (N)
 [70] Lyonissa he **saiivat** melko pian varmuuden, että he nyt olivat oikeilla jäljillä. (N)

Merkitysluokka 6: saada tulos laskemalla tai päättelemällä

Luokan 5 tapaan myös luokka 6 on hyvin spesifi merkitykseltään:

6. tuloksen syntymisestä laskemalla, päättelemällä tms. tavalla. *Tutkimuksessa saatiin yllättäviä tuloksia. Kun kaksi kerrotaan kahdella, saadaan neljä. Yhtälöstä saadaan a:n arvoksi viisi.*

Oppijansuomen aineiston molemmat merkitysluokan 6 virkkeet ovat selkeitä tapauksia (oppijansuomen aineistoesimerkit 71 ja 72). Natiivisuomen aineistossa luokan 6 esimerkkejä on enemmän, eikä niissäkään ole epäselvyyksiä (natiivisuomen aineistoesimerkit 73–76):

- [71] [– –] opettajan täytyy tehdä testi niin, että **saada** tulos, jota hän halua tietää. (O)
 [72] On aika hämmentävää, että aina, kun teen testin kuin nyt Iltalehden sivuissa, **saan** tuloksena, että olen erittäin auttavainen ihminen, joka vahtii ja suojelee muita. (O)
 [73] Mutta mitään tilastollisesti merkittävää näyttöä ei koskaan **saatu** [– –]. (N)
 [74] Ehkäisytablettien pitkäaikainen käyttö, erityisesti ennen ensimmäistä synnytystä, suurentaa joidenkin tutkimusten mukaan rintasyöpätavaraa, mutta vakuuttavaa osoitusta tästä ei ole **saatu**. (N)
 [75] Jos voidaan osoittaa, että jonkin rakenneosan historiallinen synty on yhteydessä yhteiskunnan tiettyyn rakennemuutokseen, **saadaan** vähintäänkin jotain empiiristä perustaa sen funktioiden pohtimiseksi. (N)
 [76] Jos lähtökohtana pidetään sitä, että yksi henkilö tarvitsi 70 km²:n alueen, **saadaan** kivikauden Suomeen asukkaita 5000–6000. (N)

Merkitysluokka 7: saada asema tai olotila

Luokassa 7 SAADA-verbin merkitys on ‘jonkin aseman, olotilan, tehtävän tms. tulemisesta jonkin osaksi; saavuttaa, tavoittaa’. Kielitoimiston sanakirja antaa merkitykselle 7 seuraavat esimerkit:

Muotokuva sai huoneessa kunniapaikan. Maa sai itsenäisyyden. Tuli sai (yli)vallan. Saada jalansijaa. Ei saanut lapsilta hetkenkään rauhaa. Saada rauha sielulle, tunnolle. Saada anteeksi. Ei saanut unta *ei kyennyt nukahtamaan*. Saada jstak kyllikseen, kylläkseen, tarpeekseen *kyllästyä jhk*. Sai asian toimittakseen, toimitettavakseen. Sai muodostaakseen hallituksen. Toimekseen *t. toimeksi saaneena* N. [*asiakirjan allekirjoituksena*].

Seuraavaksi merkitysluokkaa 7 havainnollistetaan nostamalla esimerkeiksi tyypillisiä luokan 7 virkkeitä (oppijansuomen aineistoesimerkit 77–83 ja natiivisuomen aineistoesimerkit 84–87). Oppijansuomen aineistoesimerkin 83 huomautettakoon, että Jussi-ehdokkuus mielletään asemana; Jussi-palkinto sen sijaan on jonkinlainen ‘lahja’, minkä vuoksi merkitys ‘saada Jussi-palkinto’ on luokiteltu luokkaan 1.

- [77] Hänen tilanne oli vielä huonompi siksi että hän ei **saanut** unta ollenkaan! (O)
 [78] Ensinnäkin – eihän ketään voi noin vain häätää, kun hän on kerran **saanut** turvapaikan! (O)
 [79] Vuodesta 2005 kveeni **sai** vähemmistökielen statusta. (O)
 [80] Lukion aikana **saimme** monia tehtäviä jossa meidän täytyi panna sanoihin mitä näemme kuvassa. (O)
 [81] Baby halusi **saada** anteeksi, että hän oli ottanut tabletteja. (O)
 [82] Molemmat olivat **saaneet** tarpeekseen. (O)
 [83] Elokuva **sai** yksitoista Jussi-ehdokkuutta ja myös Suomen Oscarin-ehdokkaan. (O)
 [84] [– –] pienen kansan äidinkielet **saavat** oikeutuksen kansainvälisessä yhteistyössä. (N)
 [85] Pääjohtaja af Enehjelm [– –] **sai** myös senaatin vahvistuksen. (N)
 [86] Helpompi olisi ollut **saada** oikaisu Helsingin Sanomiin. (N)
 [87] Ylämaan kaupunkien piti puolestaan **saada** yksinoikeus kauppaan seutujensa maalaisväestön kanssa. (N)

Luokka 7 on melko moninainen, eikä siis kovin selvärajainen, ja luokkien 3 ja 7 välillä on jouduttu käymään paljon rajanvetoa (oppijansuomen aineistoesimerkki 88, natiivisuomen aineistoesimerkki 89).

[88] Se **sai** ensi-iltansa vuonna 2005. (O)

[89] Vaiennetut voivat **saada** äänen, aiemmin kulttuurin ulkopuolelle jääneet kokemukset ja tietoisuus tulla esiin, ja pilkkanimet saada uuden merkityksen pilkattujen ottaessa ne käyttöönsä (kuten sana 'nigger' nuorten yhdysvaltalaisen mustien itsestään nykyään käyttämänä nimityksenä). (N)

Oppijansuomen aineistoesimerkin 88 *ensi-illan* voisi mieltää myös niin, että se ikään kuin tulisi elokuvan osaksi. Ensi-ilta nähdään kuitenkin elokuvan (kuten myös edellä mainitun Jussi-ehdokkuuden) väliaikaisena tilana ja asemana, minkä vuoksi tässä tapauksessa päädyttiin luokkaan 7. Natiivisuomen aineistoesimerkin 89 *äänen* voisi tulkita myös ilmaantuvan vaiennettujen osaksi, jolloin virke kävisi luokkaan 3. Koko kontekstin valossa tulkitaan kuitenkin, että tässä kontekstissa 'ääni' tarkoittaa valta-aseman saamista.

Merkitysluokka 8: saada jotakin tehdyksi tai saada joku tekemään jotain

Luokan 8 merkitys on: 'onnistua, kyetä tekemään tai aiheuttamaan jotakin tai saattamaan tekeminen päätökseen' (KS), johon Kielitoimiston sanakirja on antanut seuraavat esimerkit:

Korjaamo sai autosta vielä käyttökelpoisen. Luista saatiin hyvä keitto. Lapsia oli vaikea saada vuoteeseen. Ei saanut asiaa mielestään. Etkö saa sanaa suustasi! Sai talonsa vesikattoon. Tehtyä ei saa tekemättömäksi. Saada jku hengiltä. Jännitys sai hänet vapisemaan. Saada jtk aikaan. Sain vihdoon kirjeen aikaan Sinulle. Saada jtk aikaiseksi *ark. aikaan*. Saada jk esiin, irti, auki, rikki. Sai työn tehdyksi t. tehtyä. Vahingot saatiin korjatuksi t. korjattua. Sai kirjoituksensa julkaistuksi.

Molemmissa aineistoissa oli runsaasti luokan 8 esiintymiä (oppijansuomen aineistoesimerkit 90–93 ja natiivisuomen aineistoesimerkit 94–97):

[90] En tietenkään ole **saanut** mitään tehdyksi, onneksi on vielä aikaa jäljellä. (O)

[91] Minna Karlsson-Kananen **saa** miehiä kunnioittaa naisia [– –]. (O)

[92] Liikunta todella vaikuttaa tunnetilaan, piristää ja **saa** minut hyvälle tuulelle. (O)

[93] [– –] ne piilottavat niin syvällä sohvan alla, etten voinut **saada** niitä sieltä pois ja takaisin häkkiin. (O)

[94] Haaveilen **saavani** edes yhden puuni kasvamaan niin, että voin istuttaa sen kerrostalomme pihaan. (N)

[95] [– –] eikä kaikkia Räikkösiä ole **saatu** tapettua tai linnaan, koska pitää kyytiä tilata. (N)

[96] Tästä syystä haetaan usein rauhoittavia lääkkeitä, jotta nämä pelottavat tunteet **saataisiin** kuriin. (N)

[97] Erilaiset ympäristömuutokset ja niiden uhkat **saavat** aikaan stressiä, jonka tiedetään johtavan elimellisiin ja sosiaalisiin haittoihin. (N)

Luokka 8 on melko selväraajainen, vaikka pientä päällekkäisyyttä saattaa esiintyä luokan 4 kanssa. Kuten esimerkeistä huomaa, luokassa 8 tyypillistä on, että joku saa tehtyä jotain, tai joku saa jonkun tekemään jotain.

Merkitysluokka 9: saada lupa

Kielitoimiston sanakirja määrittelee luokan 9 esimerkkeineen seuraavasti:

9. ilmaisemassa että jk on luvallista, sallittua. *Saanko lainata kynäsi? Sai pitää löytämänsä esineen. Tili, jolta saa nostaa varoja rajoituksetta. Lämmityslaitetta ei saapeittää. Älä, ei saa lyödä!*
Erik. kohteliaisuutta ilmaisemassa. *Saanko tarjota lisää kahvia? Mitä Teille saisi olla [= mitä haluaisitte ostaa t. ottaa]? Saanen vielä korostaa yhtä seikkaa. Pyydänsaada huomauttaa, että –.*

Luokan 9 merkitys on periaatteessa tarkkarajainen tai selvä, mutta luokissa 9, 10, 11, 12 ja 13 alkaa korostua virkkeen aspekti, puhujan suhtautuminen asiaan. Näissä luokissa on monia virkeitä, joiden luokittelumiseksi on täytynyt tarkistaa laajempi konteksti suhtautumisen ja sävyn selvittämiseksi. Seuraavaksi esittelen selkeitä tapauksia luokasta 9 (oppijansuomen aineistoesimerkit 98–100 ja natiivisuomen aineistoesimerkit 100–103):

- [98] Suomessa ei **saa** rangaista lapsia [– –]. (O)
- [99] Minusta maa, jossa ihminen ei **saa** sanoa mitään valtaa vastaan ja pelkää elämästään joka päivä, ei yksinkertaisesti voi olla hyvä ajatus. (O)
- [100] Jokainen **saa** sanoa oman mielipiteensä. (O)
- [101] Avioeroa riidellessään mies puhuu autosta – kuka sillä saa ajaa ja minne. (N)
- [102] Kaikenlaistahan **saa** kuvitella! (N)
- [103] Etikkaa ei **saa** sekoittaa hunajaan. (N)

Seuraavaksi esittelen tapauksia, jotka eivät ole aivan yhtä selkeitä, mutta jotka on kuitenkin luokiteltu merkitykseen 9. Näissä tapauksessa SAADA-verbi voisi merkitä myös mahdollisuutta johonkin, mutta sävy tuntuu viittaavan enemmän lupaan kuin mahdollisuuteen (oppijansuomen aineistoesimerkit 104–105 ja natiivisuomen aineistoesimerkit 106–109). Esimerkkien 104–109 laajemmat kontekstit on tarkistettu tarkan sävyn selvittämiseksi.

- [104] Onni että **saimme** tulla hänen kyydissä. (O)
- [105] Kissa on erittäin tyytyväinen, kun se **sai** syödä niin paljon tänä iltana. (O)
- [106] Jos Heikki on kuollut, miksi minä en **saa** nähdä? (N)
- [107] Avasin Pöyhösen kämpän oven ja sanoin liikalle, että se **saisi** olla siellä vähän aikaa. (N)
- [108] Pussissa oli monta omenaa, ja minulle sanottiin, että **saisin** yksin pitää koko pussin sisällön. (N)
- [109] Alakerrassa tunnelma oli ennallaan, mutta Katja oli **saanut** muuttaa huoneensa nykyaikaiseksi. (N)

Luokan 9 erikoismerkitys on kohteliaisuuden ilmaiseminen, esimerkiksi: *Mitä teille saisi olla?* Erikoismerkitystä havainnollistavat oppijansuomen aineistoesimerkki 110 ja

natiivisuomen aineistoesimerkit 111–113, joista viimeinen on luokassa 9 korostetun kohteliaisuuden luoman sarkastisen sävyn vuoksi: ilman sitä se kuuluisi luokkaan 1b ‘saada lupa’.

[110] **Saisinko** tarjota sinulle illallisen ravintola Kellarissa? (O)

[111] Teinä, jos **saan** sanoa, käyttäisin viimeiset kuukauteni johonkin arvokkaampaan kuin kuolemanpelosta nauttimiseen, viimeisistä hetkistäni hekumoimiseen. (N)

[112] **Saanko** nyt kysyä erästä hieman intiimimpää asiaa. (N)

[113] Ja mitä tekemään, jos **saa** luvan kysyä? (N)

Merkitysluokka 10: saada tapahtua tai olla

Luokan 10 merkitys on, että ’jokin sopii, ei tuota haittaa, on yhdentekevää’, mikä tosin kielteisenä tarkoittaa, että ’jokin ei sovi, tuottaa haittaa tai ei ole yhdentekevää’, kuten KS:n viimeisestä esimerkeistä selviää:

Sato on korjattu, talvi saa tulla. Ellet usko, saat olla uskomatta. Nyt en saa epäonnistua.

Oppijansuomessa merkitystä 10 on vain yksi tapaus (esimerkki 114). Natiivisuomen aineistossa tapauksia on useampia, joista esimerkeiksi on valikoitu edustavimmat (esimerkit 115–117). Kuten luokan 9 yhteydessä tuotiin esille, myös luokassa 10 luokitukseen vaikuttaa hyvin paljon pienet nyanssit, jotka ohjaavat tulkitsemaan suhtautumista asiaan.

[114] [--] ’jalat **saa** olla tikut, persettä ei telttahousuista kuka erota’ [--]. (O)

[115] Todennäköisyytenä tämä ei kuitenkaan **saa** ennakoida tulevaan, koska eettisestä pyrkimyksestä luopuminen ei tunne mitään rajaa. (N)

[116] [--] miespuolisille vieraille tarjoillaan väkijuomia ja tämä **saa** tapahtua kaikessa rauhassa. (N)

[117] Kukaan ei **saisi** kuulla Lauran olevan valveilla. (N)

Merkitysluokka 11: tilaisuus, mahdollisuus

Luokassa 11 SAADA-verbi ilmaisee ’tilaisuutta, mahdollisuutta jhk’ tai on merkityksessä ’voida, saattaa’. Esimerkkeinä KS antaa seuraavat:

Tilatut tavarat saa noutaa huomenna. Taikinan on saatava nousta kunnolla. Täällä saamme olla rauhassa. Ei saanut melulta t. melun vuoksi nukutuksi t. nukuttua.

’Tilaisuus’ tai ’mahdollisuus’ johonkin pitävät nähtävästi sisällään SAADA-verbin merkityksen ’päästä tekemään jotakin’. Tämä merkitys lienee lähellä luokan merkitystä ’voida’. (Oppijansuomen aineistoesimerkki 118 ja natiivisuomen aineistoesimerkki 119.)

[118] Kesäkin on mukava aika, koska silloin **saa** ottaa aurinkoa ja uida sekä matkustaa ja lomailla. (O)

[119] Minä tarvitsen ja haluan että minulla on ihminen, jota **saan** rakastaa. (N)

Vaikka luokassa 11 SAADA-verbi merkitsee 'tilaisuutta', 'saada tilaisuus johonkin' kuuluu KS:n omien esimerkkien mukaan kuitenkin luokkaan 1b. Jälleen kyse on siis syvemmästä merkityksestä kuin pelkästä sanatasosta.

Merkitysluokka 12: pakollisuus ja väistämättömyys

Luokan 12 merkitys on: 'ilmaisemassa jnk pakollisuutta t. väistämättömyyttä; joutua tekemään jtk'. Luokasta annetaan seuraavia esimerkkejä:

Tämän hän saa vielä maksaa. Sai tehdä paljon turhaa työtä. Sai nuolla näppejään. Tästä saat syyttää itseäsi. Saat uskoa, että kaikki vielä järjestyy. Aina häntä saa odottaa. Parempaa miestä saa hakea *on vaikea löytää*.

Luokat 11 ja 12 ovat hyvin samankaltaisia, ja niitä erottaa lähinnä virkkeen sävy. Luokassa 11 sävy on positiivinen, luokassa 12 negatiivinen. Esimerkki 120 voisi olla joko mahdollisuutta johonkin niin, että kyseessä olisi iloinen asia, tai pakollisuutta johonkin niin, että verbi on merkityksessä 'joutua'. Laajemmasta kontekstista selvisi, että esimerkeissä 120 kyseessä on nimenomaan 'joutuminen', jonkinlainen 'pakollisuus', ja esimerkissä 121 'väistämättömyys'.

[120] Aino **saa** olla tyytyväinen kun hän saa uuden tyynyliinan missä on sisiliskon kuva. (O)

[121] Häntä **saan** kiittää, miten vähin vaurioin sittenkin aina kolhin itseni kotiin. (N)

Merkitysluokka 13: kehotus ja toivomus

Luokassa 13 SAADA-verbi ilmaisee (moitteeseen liittyvää) käskyä, kehotusta tai toivomusta:

Nyt saa riittää! Ellet tottele, saat jäädä ilman jälkiruokaa. Saat luvan pyytää anteeksi! Saisitte pitää lapsenne paremmassa kurissa. Verotus saisi olla lievempää.

Oppijansuomen aineistossa luokasta 13 on vain kaksi esiintymää (oppijansuomen aineistoesimerkit 122–123). Natiiviaineistossa esimerkkejä on runsaammin, joista tähän on valittu edustavimmat (natiivisuomen aineistoesimerkit 124–127):

[122] **Saat** syyttää vain itsesi. (O)

[123] Ajattelin jo, että voitaisiin hankkia marsu – se **saisi** syödä ne. (O)

[124] [– –] Olisi misterit tosiaan **saanut** antaa Gettalle tämän illan vapaata. (N)

[125] Jos se ei jollekin sovi niin työpaikkaa **saa** vaihtaa. (N)

[126] Enempikin **sais** mun mielestä osallistua. (N)

[127] Jos möykkääminen ei loppu tällä alhaalla niin te **saatte** lähteä molemmat. (N)

Luokassa 13 on hyvä esimerkki siitä, kuinka virkkeen sävy vaikuttaa tulkintaan (natiivisuomen aineistoesimerkki 128). Mikäli virkettä tarkastelisi pelkästään sanatasolla, ‘saada lupa’ menisi se luokkaan 1b.

[128] Susi saisi luvan nostaa hattuaan ja kiertää hänet kaukaa. (N)

Merkitysluokka 14: saada nähdä ja merkitysluokka 15: vanhahtava

Luokassa 14 SAADA-verbi esiintyy NÄHDÄ-verbin ohella tulevaisuuteen viittaavissa ilmauksissa (KS s. v. *saada*). Tästä luokasta on vain yksi natiivisuomen aineistoesimerkki 129:

[129] **Saan** nähdä, Janne sanoo. (N)

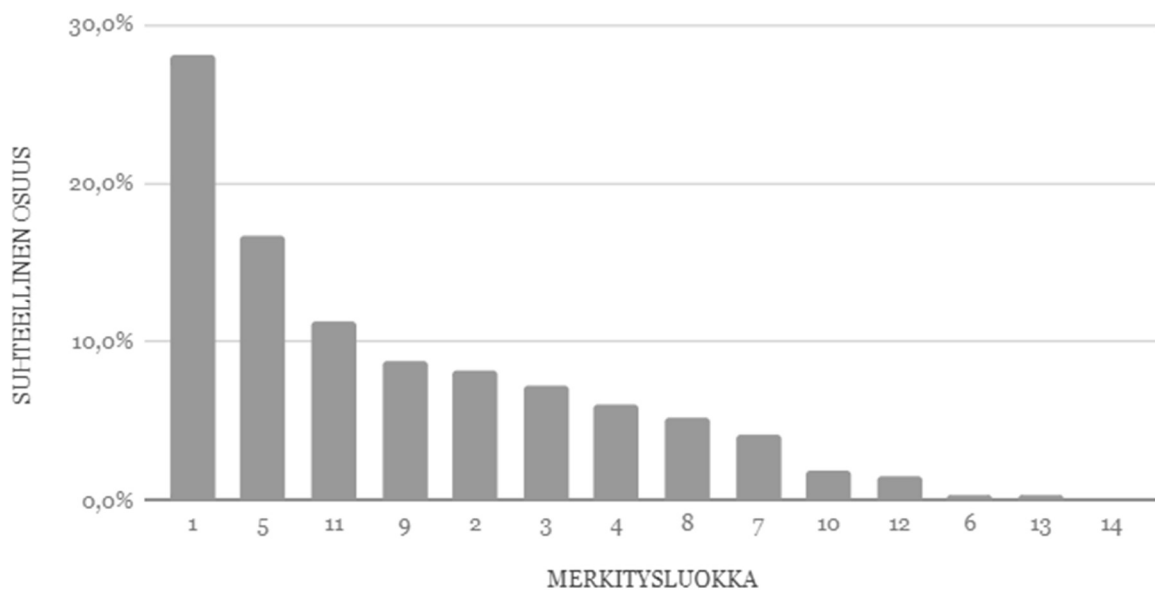
KS:n luokituksessa on myös luokka 15, jossa SAADA-verbi määritellään vanhahtavaksi ja runokieliseksi. Tätä luokkaa ei esiintynyt kummassakaan aineistossa, joten se jätetään käsittelemättä tarkemmin.

6.1.2 SAADA-verbin merkitykset oppijan- ja natiivisuomessa

Tässä alaluvussa vastataan ensimmäiseen tutkimuskysymykseen. Oppijansuomen aineiston (513 virkettä) jakautuminen merkitysluokkiin on kuvattu kuviossa 4 (ks. tarkempi taulukko liite 2). Suurin merkitysluokka oppijansuomessa on 1 ‘tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkun tai jonkin omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kanalta, jonka omistukseen, käyttöön tms. jokin tulee’, ja sen osuus koko aineistosta on lähes kolmasosa, 28,1 % (n = 144). Seuraavaksi eniten on luokkaa 5 ‘jonkin tulemisesta jonkun aistimusten, tiedon tai ymmärryksen piiriin’, 16,8 % (n = 86). Kolmanneksi taajin on merkitysluokka 11 ‘ilmaisemassa tilaisuutta, mahdollisuutta johonkin; voida, saattaa’, 11,3 % (n = 58). Huomattavaa on, että nämä kolme taajinta merkitysluokkaa kattavat yhteensä yli puolet (56,2 %, n = 288) aineistosta.

Loput esiintymät jakautuvat taajuusjärjestyksessä merkitysluokille 9 ‘ilmaisemassa että jokin on luvallista, sallittua’ (8,8 %, n = 45), 2 ‘tulla, joutua jonkin toimenpiteen, käsittelyn, kohtelun tms. kohteeksi’ (8,2 %, n = 42), 3 ‘ilmaantua tai tulla jonkin osaksi tai hyväksi’ (7,2

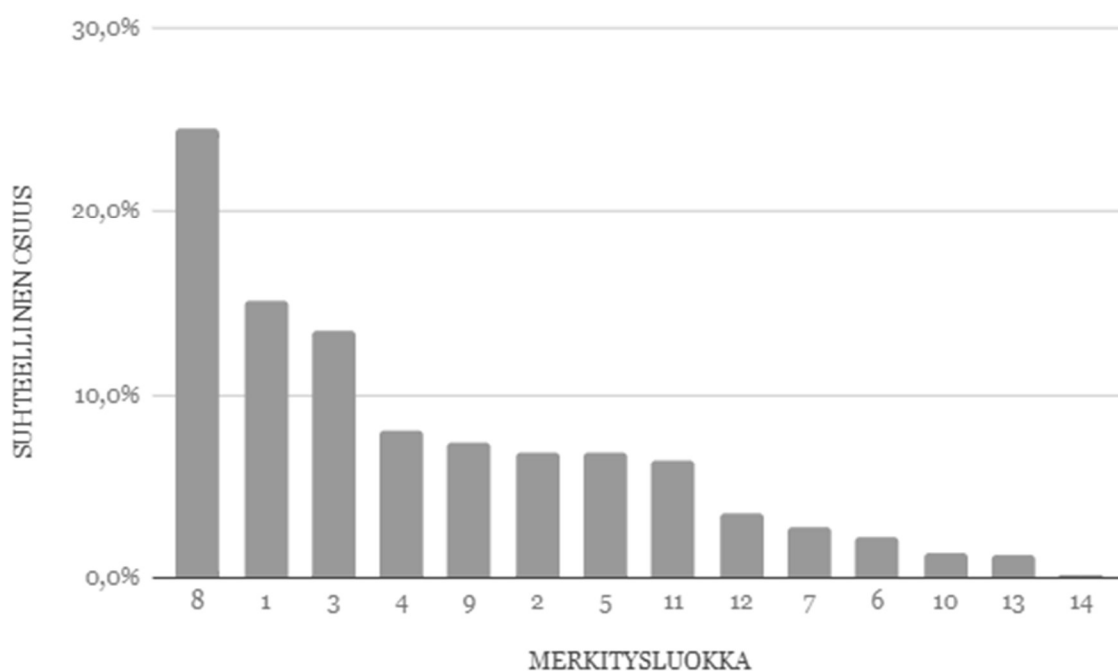
%, n = 37), 4 ‘onnistua hankkimaan jotakin itselleen, käyttöönsä, yhteyteensä tms., tavoittaa, saavuttaa’ (6,0 %, n = 31), 8 ‘onnistua, kyetä tekemään tai aiheuttamaan jotakin tai saattamaan tekeminen päätökseen’ (5,3 %, n = 27), 7 ‘jonkin aseman, olotilan, tehtävän tms. tulemisesta jonkin osaksi; saavuttaa, tavoittaa’ (4,1 %, n = 21), 10 ‘ilmaisemassa, että jokin sopii, ei tuota haittaa, on yhdentekevää’ (1,9 %, n = 10), 12 ‘ilmaisemassa jonkin pakollisuutta tai väistämättömyyttä; joutua tekemään jotakin’ (1,6 %, n = 8), 6 ‘tuloksen syntymisestä laskemalla, päättelöllä tms. tavalla’ (0,4 %, n = 2) ja 13 ‘ilmaisemassa (moitteeseen liittyvää) käskyä, kehoitusta tai toivomusta’ (0,4 %, n = 2). Oppijansuomen aineistossa ei esiintynyt luokkia 14 ja 15, ja kuten edellä ilmeni, myös luokkien 6 ja 13 on melko marginaalista.



Kuvio 4. Oppijankielen aineiston jakautuminen merkitysluokkiin (513 virkettä).

Tulokset kertovat siitä, että oppijansuomessa SAADA-verbi esiintyy taajimmin sellaisessa merkityksessä, että jokin asia konkreettisesti siirtyy jollekin, jonka näkökulmasta asiaa katsotaan (esim. ‘saada esine, etuus tai rahaa’). SAADA-verbin merkitysluokan 1 semanttista ympäristöä käsitellään tarkemmin myöhemmin luvussa 6.3. Lisäksi SAADA-verbi esiintyy taajasti sellaisissa merkityksessä, jossa kyse on jonkinlaisen tiedon tai ymmärryksen, tai jonkinlaisen mahdollisuuden ja tilaisuuden saamisesta. Sen sijaan oppijansuomessa SAADA-verbi esiintyy vain harvoin sellaisissa merkityksissä, jotka koskevat (tieteellisten tai laskettujen) tulosten saamista, käskemistä, kehoitusta tai toivomista. SAADA-verbi ei esiinny (tämän tutkimuksen oppijansuomen aineistossa) lainkaan NÄHDÄ-verbin ohella tulevaisuuteen viittavissa ilmauksissa (esim. *saa nähdä*) tai vanhahtavassa, runollisessa merkityksessä (esim. ‘aamu saa’).

Natiivisuomen aineiston (813 virkettä) jakautuminen merkitysluokkiin on kuvattu kuviossa 5 (ks. tarkempi taulukko liite 3). Jakautuminen näyttäisi kuitenkin olevan hieman tasaisempaa kuin oppijansuomen aineistossa, jossa taajin merkitys kattaa lähes kolmasosan esiintymistä; natiivisuomen aineistossa taajin merkitys kattaa lähes neljäsosan (24,5 %, n = 199). Täytyy kuitenkin muistaa, että natiivisuomen aineisto on noin puolitoistakertainen oppijansuomen aineistoon nähden, joten tasaisempi jakautuminen lienee odotuksenmukaista. Natiivisuomen taajin merkitysluokka on 8 'onnistua, kyetä tekemään tai aiheuttamaan jotakin tai saattamaan tekeminen päätökseen', joka on oppijansuomen kahdeksannella sijalla. Toiseksi taajin luokka natiivisuomessa on oppijansuomen taajin luokka 1 'tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkun tai jonkin omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jokin tulee' (15,1 %, n = 123). Kolmanneksi taajin (oppijansuomessa sijalla kuusi) merkitys natiivisuomessa on 3 'ilmaantua tai tulla jonkin osaksi tai hyväksi' (13,5 %, n = 110).



Kuvio 5. Natiivisuomen aineiston jakautuminen merkitysluokkiin (813 virkettä).

Loput luokat taajuusjärjestyksessä ovat 4 'onnistua hankkimaan jotakin itselleen, käyttöönsä, yhteyteensä tms., tavoittaa, saavuttaa' (8,1 %, n = 66), 9 'ilmaisemassa että jokin on luvallista, sallittua' (7,4 %, n = 60), 2 'tulla, joutua jonkin toimenpiteen, käsittelyn, kohtelun tms. kohteeksi' (6,9 %, n = 56), 5 'jonkin tulemisesta jonkin aistimusten, tiedon tai ymmärryksen piiriin' (6,9 %, n = 56), 11 'ilmaisemassa tilaisuutta, mahdollisuutta johonkin; voida, saattaa' (6,4 %, n = 52), 12 'ilmaisemassa jonkin pakollisuutta tai väistämättömyyttä, joutua

tekemään jotakin' (3,6 %, n = 29), 7 'jonkin aseman, olotilan, tehtävän tms. tulemista jonkin osaksi; saavuttaa, tavoittaa' (2,7 %, n = 22), 6 'tuloksen syntymisestä laskemalla, päättelemällä tms. tavalla' (2,2 %, n = 18), 10 'ilmaisemassa, että jokin sopii, ei tuota haittaa, on yhdentekevää' (2,2 %, n = 18), 13 'ilmaisemassa (moitteeseen liittyvää) käskyä, kehotusta t. toivomusta' (1,2 %, n = 10) ja 14 'nähdä-verbin ohella eräissä tulevaisuuteen viittaavissa ilmauksissa' (0,1 %, n = 1). Natiivisuomessa SAADA-verbi näyttäisi siis esiintyvän sellaisissa merkityksissä, joissa kyse on jonkin asian aikaan saamisesta tai tekemisen onnistumisesta ja loppuun päättämisestä, jonkin asian konkreettisesta siirtymisestä jollekin, jonka näkökulmasta asiaa katsotaan tai jonkin asian ilmaantumisesta jonkin osaksi.

Huomattavan arvoista on, että oppijan- ja natiivisuomen kolmen taajimman merkitysluokan joukossa on vain yksi sama luokka (1 'tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkun tai jonkin omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jokin tulee'). Tämä kertoo siitä, että SAADA-verbi käyttäytyy polyseemisesti eri tavoin oppijan- ja natiivisuomessa.

Seuraavaksi tarkastellaan oppijan- ja natiivisuomen merkitysluokkien välisiä tilastollisia eroja. Näitä eroja kuvataan kuviossa 6, joka on järjestetty vasemmalta oikealle tilastollisesti merkittävimmistä eroista vähiten merkitseviin. Tilastollisen merkitsevyyden testaamiseen on käytetty z-testiä. Tarkempi taulukko merkitysluokkien z-testistä on liitteenä 4, jossa ovat myös tarkat z-arvot ja keskivirheet.

Luokkien 8 'onnistua, kyetä tekemään tai aiheuttamaan jotakin tai saattamaan tekeminen päätökseen'²¹, 1 'tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkun tai jonkin omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jokin tulee'²², 5 'jonkin tulemisesta jonkun aistimusten, tiedon tai ymmärryksen piiriin'²³ ja 3 'ilmaantua tai tulla jonkin osaksi tai hyväksi'²⁴ erot ovat tilastollisesti erittäin merkitseviä, luokkien 11 'ilmaisemassa tilaisuutta, mahdollisuutta jhk; voida, saattaa'²⁵ ja 6 'tuloksen syntymisestä laskemalla, päättelemällä tms. tavalla'²⁶ erot tilastollisesti merkitseviä ja luokka 12 'ilmaisemassa jonkin pakollisuutta tai väistämättömyyttä; joutua tekemään jotakin'²⁷ on tilastollisesti melkein merkitsevä. Tilastollisen merkitsevyyden vuoksi on paikallaan tarkastella luokkien käyttöeroja myös laadullisesti. Tutkielman laajuuden vuoksi tarkastellaan kuitenkin

²¹ z=-9,06, p<0,001.

²² z=5,72, p<0,001.

²³ z=5,66, p<0,001.

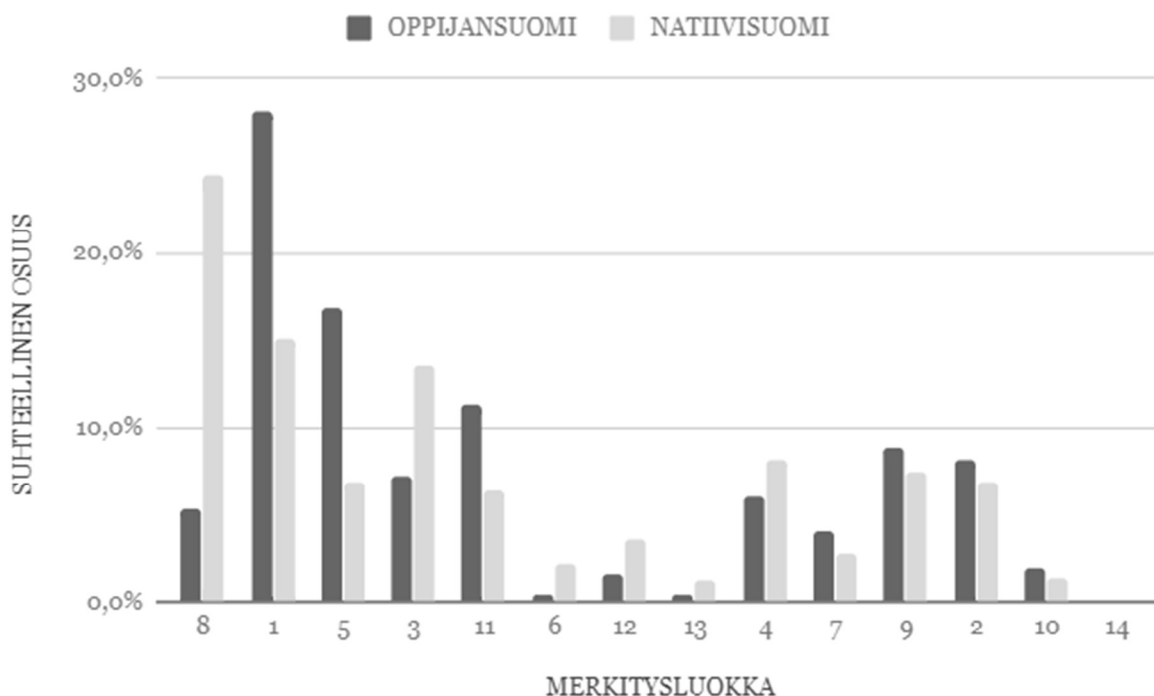
²⁴ z=-3,57, p<0,001.

²⁵ z=3,16, p<0,01.

²⁶ z=-2,65, p<0,01.

²⁷ z=-2,16, p0,05.

vain neljää luokkaa, joissa erot ovat erittäin merkitseviä (vaikka muutkin luokat olisivat relevantteja tarkastelulle laajemmassa tutkielmassa).



Kuvio 6. Oppijansuomen (513 virkettä) ja natiivisuomen (813 virkettä) merkitysluokkien vertailu järjestettynä tilastollisen merkitsevyyden²⁸ mukaan.

Merkittävin SAADA-verbin polyseeminen käyttöero on siis luokan 8 ‘onnistua, kyetä tekemään tai aiheuttamaan jotakin tai saattamaan tekeminen päätökseen’ suhteen siten, että natiivisuomessa sitä käytetään huomattavan paljon enemmän kuin oppijansuomessa, eli oppijansuomessa tämä merkitys aliedustuu. Oppijansuomessa luokassa 8 tuntuu korostuvan merkitys ‘saada joku tekemään jotain’ (oppijansuomen esimerkit 130–133). Ainakin osa tällaisista rakenteista on peräisin samasta tehtävänannosta, jossa venäjänkielisten oppijoiden on täytynyt analysoida novellia (oppijansuomen esimerkit 131–132). Voisi siis myös pohtia, onko tälle kyseiselle kurssille opetettu tämä rakenne erikseen. Toisaalta, koska luokka 8 on kuitenkin taaempi natiivisuomessa, tällainen yhden tehtävän poikkeama oppijansuomessa vahvistaa käytön eroa (koska ilman tämän oppijansuomen tehtävän aineistoa oppijansuomen luokan 8 osuus olisi vielä pienempi, mikä lisäisi eroa entisestään natiivisuomeen).

[130] [–] vuonna 2000 **saatiin** Mika Häkkinen ja David Coulthard vierailemaan Prahassa. (O)

[131] Ajattelen, että isäpuoli toimii ärsykkeenä tässä novellissa ja **saa** pojan miettimään ja tuntemaan. (O)

²⁸ Raja-arvot: 1,96 melkein merkitsevä ($p < 0,05$); 2,57 merkitsevä ($p < 0,01$); 3,29 erittäin merkitsevä ($p < 0,001$). Arvot on pyöristetty kahden desimaalin tarkkuudelle.

- [132] [– –] ja jopa se pieni katkelma voi **saada** lukijoita nauramaan ja olemaan hyvällä tuulella. (O)
 [133] [– –] palkan lisääminen **saisi** heidät jatkamaan työssä entistä pitempään. (O)

Natiivisuomessa sen sijaan tuntuu korostuvan merkitykset 'saada jotain tehdyksi' tai 'onnistua tekemään jotain' (natiivisuomen esimerkit 134–137). Natiivisuomen aineistosta ei ole samalla tavalla havaittavissa, että tietäntyyppiset rakenteet tulisivat samasta aihepiiristä (tai samasta tekstistä).

- [134] Sen sijaan **saat** nauhoitettua äänten kakofonian, äänten kaaoksen. (N)
 [135] [– –] arveltiin niiden yöllä yrittävän talon valtausta [– –], jotta Jeltsin **saataisiin** syrjään. (N)
 [136] [– –] murtumakohta joudutaan murtamaan uudelleen, jotta luut **saataisiin** oikeaan asentoon. (N)
 [137] [– –] kaikkia Räikkösiä ole **saatu** tapettua tai linnaan, koska pitää kyytiä tilata. (N)

Luokan 1 'tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkun tai jonkin omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jokin tulee' käyttöero on myös tilastollisesti erittäin merkitsevä niin päin, että oppijansuomen aineistossa merkitys yliedustuu. Oppijansuomen aineistosta erottuu natiivisuomeen nähden merkitys 'saada lahja' (oppijansuomen aineistoesimerkit 138–140). Merkityksen 'saada lahja' korostuminen näkyy myöhemmin myös semanttisen primingin analyysissä, joka on tehty oppijansuomen taajimman merkityksestä, eli luokasta 1, kun taas natiivisuomessa korostuvat muut konkreettiset entiteetit (natiivisuomen esimerkit 141–143):

- [138] Jos olet vihainen et **saa** joululahjoja Joulupukista! (O)
 [139] On kiva **saada** lahjoja [– –]. (O)
 [140] Sitten **saame** lahjoja. (O)
 [141] [– –] Helsingin ja Turun tullikamarit **saivat** toiset partioveneensä. (N)
 [142] **Saat** tämän kasetin itsellesi, Topi sanoi. (N)
 [143] [– –] jos **sais** parttion poroja ennen uutta kuuta. (N)

Luokassa 5 'jonkin tulemisesta jonkun aistimusten, tiedon tai ymmärryksen piiriin', kuten myös edellä käsitellyssä luokassa 1, tilastollinen ero aineistojen suhteellisissa osuuksissa on niin päin, että luokkaa käytetään merkittävästi enemmän oppijansuomessa kuin natiivisuomessa. Oppijansuomen aineistossa luokan 5 tapaukset liittyvät vahvasti merkitykseen 'saada tietää' (oppijansuomen aineistoesimerkit 144–146), kun taas natiivisuomessa (natiivisuomen aineistoesimerkit 147–149) virkkeet varioivat enemmän:

- [144] Novellin loppu on avain, me emme **saa** tietää 'kuinka henkilöyden käy'. (O)
 [145] Kun minä opin, minun täytyy ensiksi **saada** sen perustietoa ja muistaa ja ymmärtää niin paljon kun mahdollista, esm. luen paljon kirjaa ja materiaalia. (O)
 [146] Juodessani olutta kapakassa keskustelin paljon ihmisten kanssa **saadakseni** tietää, mitä he itse ajattelevat asiasta. (O)
 [147] Liikkuva ja tunnusteleva subjekti **saa** selville kohteestaan useampia puolia kuin se [– –]. (N)
 [148] En **saa** niistä tolkkua. (N)
 [149] [– –], etten **saa** melodiasta kiinni, klassista standardia. (N)

Kuten aiemmin käsitelty luokka 8, myös luokka 3 'ilmaantua tai tulla jonkin osaksi tai hyväksi' esiintyy merkittävästi useammin natiivisuomessa kuin oppijansuomessa, eli kyse on aliedustumisesta oppijansuomessa. Erityisesti erikoismerkitystä 3b 'sairauteen, oireeseen t. vammaan liittyen' käytetään natiiviaineistossa huomattavan paljon useammin kuin oppijansuomessa (natiivisuomen aineistoesimerkit 150–152). Natiivisuomessa luokan 3 virkkeistä (n = 45) yli puolet (53,3 %, n = 24) liittyvät sairauteen tai vammaan, kun taas oppijansuomessa tällaisia virkkeitä on vain yksi (2,7 %). Sen sijaan oppijansuomessa tuntuu korostuvan 'kokemusten, tunteiden ja elämysten saaminen' (oppijansuomen esimerkit 153–155). Sairauteen liittyvien virkkeiden vähyys oppijansuomen aineistossa voi johtua yksinkertaisesti siitä, ettei terveysasioita ole käsitelty tehtävänannoissa – vastaavasti natiiviaineistossa sairaustapausten korostuminen voi johtua aineistoon osin päätyneestä teoksesta, joka käsittelee poikkeuksellisen paljon terveysasioita. Oppijansuomen kokemuksellisuus ja tunteet luokassa 3 voivat johtua siitä, että voi tuntua yksinkertaisemmalta sanoa 'saada kokemuksia', 'saada jokin tunne' tai 'saada elämyksiä (saada pelko)' samanlaisilla rakenteilla kuin ilmaista ne kenties natiiville varioivampaan tapaan 'hankkia kokemuksia', 'minua alkoi pelottamaan / minusta tuli pelokas' ja 'kokea elämyksiä'²⁹.

[150] Helikobakteerin kantajista 10-20 % **saa** [– –] infektiioon liittyen peptisen haavataudin. (N)

[151] Kolme [– –] HIV-tartunnan saaneiden äitien lapsesta on **saanut** tartunnan äidistään. (N)

[152] Se väittää, että iloiset ihmiset eivät **saa** syöpää. (N)

[153] Luotan että **saan** usea uusia kokemuksia ja opin paljon lääkäriksi olemisesta. (O)

[154] Kielen oppimisesta on **saanut** pelko, että kaikki kielet ovatkin jotain niin mystisen vaikeaa [– –]. (O)

[155] Vaikka minulla ei ole poikaystävä tänä vuonna, elämä on kaunista ja olen **saanut** paljon. (O)

Loppujen luokkien suhteellisten osuuksien erot eivät ole tilastollisesti merkitseviä tai edes melkein merkitseviä³⁰. Tämä tarkoittaa sitä, että SAADA-verbin käyttö merkitysluokkien 13 'ilmaisemassa (moitteeseen liittyvää) käskyä, kehotusta tai toivomusta', 4 'onnistua hankkimaan jotakin itselleen, käyttöönsä, yhteyteensä tms., tavoittaa, saavuttaa', 7 'jonkin aseman, olotilan, tehtävän tms. tulemisesta jonkin osaksi; saavuttaa, tavoittaa', 9 'ilmaisemassa että jokin on luvallista, sallittua', 2 'tulla, joutua jonkin toimenpiteen, käsittelyn, kohtelun tms. kohteeksi', 14 NÄHDÄ-verbin ohella eräissä tulevaisuuteen viittaavissa ilmauksissa' ja 15 'vanh., runok.' käyttö (tai käyttämättömyys) ovat natiivi- ja oppijansuomessa samankaltaisia. Etenkin luokan 4 osalta tämä on mielenkiintoista, sillä se on hyvin samankaltainen kuin luokka 1, jossa käyttöero on tilastollisesti erittäin merkitsevä.

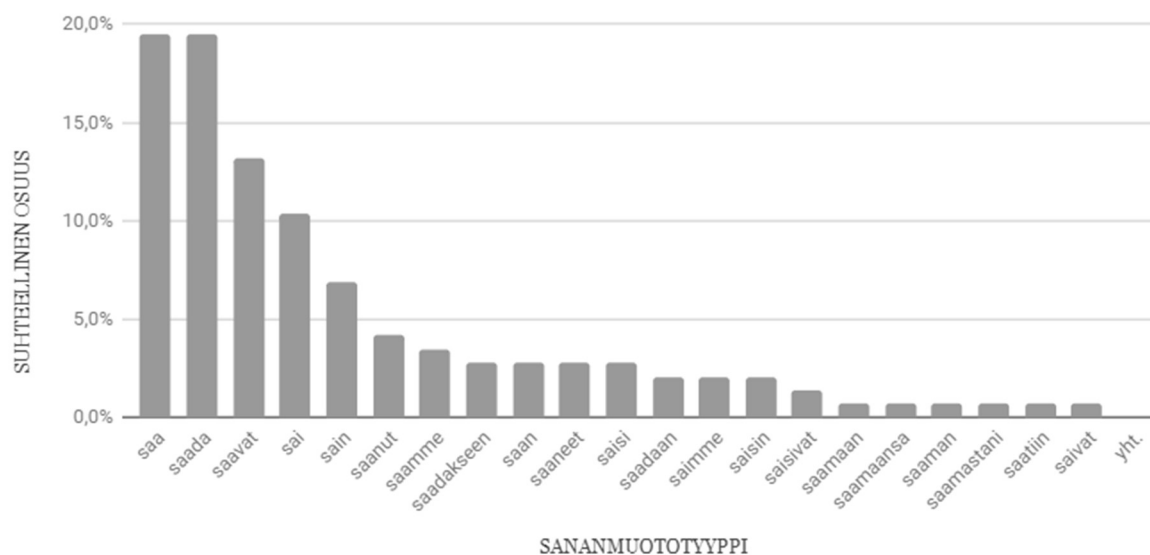
²⁹ Nämä pohdinnat perustuvat tutkijan omalle intuitiolle, eivät autenttiseen dataan.

³⁰ z=-1,96–1,96, p>0,05.

6.2 SAADA-verbin morfologinen priming oppijan- ja natiivisuomessa

Tässä luvussa vastataan toiseen tutkimuskysymykseen eli analysoidaan SAADA-verbin morfologista primingia oppijan- ja natiivisuomessa, kun SAADA-verbi on oppijansuomen aineiston taajimmassa merkityksessä 1 ‘tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkun tai jonkin omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jokin tulee’. Kuviossa 7 (ks. tarkempi taulukko liite 5) on esitetty oppijansuomen aineistossa esiintyneet SAADA-verbin 21 sananmuototyyppiä.

Oppijansuomessa (144 SAADA-verbin esiintymää) taajimmat sananmuodot ovat siis *saa* (19,4 %, n = 28) ja *saada* (19,4 %, n = 28). Loput useammin kuin kerran esiintyvät sananmuodot ovat *saavat* (13,2 %, n = 19), *sai* (10,4 %, n = 15), *sain* (6,9 %, n = 10), *saanut* (4,2 %, n = 6), *saamme* (3,5 %, n = 5), *saadakse* (2,8 %, n = 4), *saan* (2,8 %, n = 4), *saaneet* (2,8 %, n = 4), *saisi* (2,8 %, n = 4), *saadaan* (2,1 %, n = 3), *saimme* (2,1 %, n = 3), *saisin* (2,1 %, n = 3) ja *saisivat* (1,4 %, n = 2). Loput sananmuodot esiintyvät vain kerran (0,7 % per sananmuototyyppi): *saamaan*, *saamaansa*, *saaman*, *saamastani*, *saatiin* ja *saivat* (0,7 %, n = 1). Kuten merkitysluokissa, myös sananmuototyypeissä kolme taajinta esiintymää kattaa yli puolet kaikista esiintymistä (52,0 %, n = 75).



Kuvio 7. SAADA-verbin sananmuodot (n = 144) oppijansuomen aineiston merkitysluokasta 1.

Tarkemmin tarkasteltuna oppijansuomen aineistossa esiintyvät 21 sananmuototyyppiä (144 SAADA-verbin esiintymää) jakautuvat niin, että 61,9 % (n = 13) tyypeistä on finiittisiä (*saan*, *saa*, *saamme*, *saavat*, *sain*, *sai*, *saimme*, *saivat*, *saatiin*, *saisin*, *saisi*, *saisivat*, *saadaan*)

ja 38,1 % (n = 8) infiniittisiä (*saada, saadakseen, saaman, saamastani, saamaan, saamaansa, saanut, saaneet*). Kolmasosa (33,3 %, n = 7) sananmuototyypeistä viittaa kolmanteen persoonaan (finiittisesti yksikön ja monikon aktiivissa: *saa, saavat, sai, saimme, saisi, saisivat* ja infiniittisesti possessiivisuffiksilla: *saadakseen, saamaansa*) ja 28,6 % (n = 6) ensimmäiseen persoonaan (finiittisesti yksikön ja monikon aktiivissa: *saan, saamme, sain, saimme, saisin* ja infiniittisesti possessiivisuffiksilla: *saamastani*). Passiivi esiintyy sekä preesensissä (*saadaan*) että imperfektissä (*saatiin*). Oppijansuomen aineistossa toista persoonaa ei esiinny lainkaan. Finiittistä sananmuototyypeistä (n = 13) preesensia on 61,5 % (n = 8) (*saan, saa, saamme, saavat, saadaan, saisin, saisi, saisivat*) ja imperfektiä 38,5 % (n = 5) (*sain, sai, saimme, saivat, saatiin*). Tämä selittynee sillä, että konditionaalilla ei ole imperfektiä. Tässä kohti huomautetaan, että imperatiivia ja potentiaalia ei moduksista esiinny lainkaan. Indikatiivissa esiintyy 76,9 % finiidististä sananmuototyypeistä (mikä tarkoittaa luonnollisesti sitä, että konditionaalisia esiintymiä on 23,1 % finiidististä sananmuototyypeistä). Kuten myös preesensin yleisyyteen imperfektiin nähden, tähänkin todennäköisesti vaikuttaa konditionaalin imperfektittömyys. Tempuksista perfektiä ja pluskvamperfektiä ei tarkastella, koska ne vaativat OLLA-apuverbin: tässä tutkimuksessa huomioon on otettu SAADA-verbin sisäinen morfologinen muoto yksistään ilman verbiliittoja. Perfektiin ja pluskvamperfektiin liittyviä sananmuototyyppejä on kuitenkin finiidisten esiintymien puolella (*saanut ja saaneet*). Infiniittiset tyypit jakautuvat A-infinitiiviin (I infiniitiivi: *saada, saadakseen*), MA-infinitiiviin (III infiniitiivi: *saaman, saamastani, saamaan, saamaansa*) ja NUT-partisiippiin (II partisiippi: *saanut, saaneet*).

Oppijansuomen aineiston perusteella voisi siis sanoa, että oppijansuomen morfologinen priming SAADA-verbille on *saa* (oppijansuomen aineistoesimerkit 156–157 ja *saada* oppijansuomen aineistoesimerkit 158–159). Oppijansuomelle lisäksi näyttäisi olevan todennäköisempää, että sananmuototyyppi on finiittinen kuin infiniittinen (ero tosin ei ole tilastollisesti merkitsevä³¹). Lisäksi näyttäisi siltä, että finiittiset sananmuototyypit suosivat ensimmäistä ja kolmatta persoonaa ja välttelevät toista persoonaa. Infiniittisistä tyypeistä puolet (50,0 %, n = 4) on MA-infinitiivejä, minkä voisi myös mieltää jonkinasteiseksi primingiksi.

[156] Jos olet vihainen et **saa** joululahjoja Joulupukista! (O)

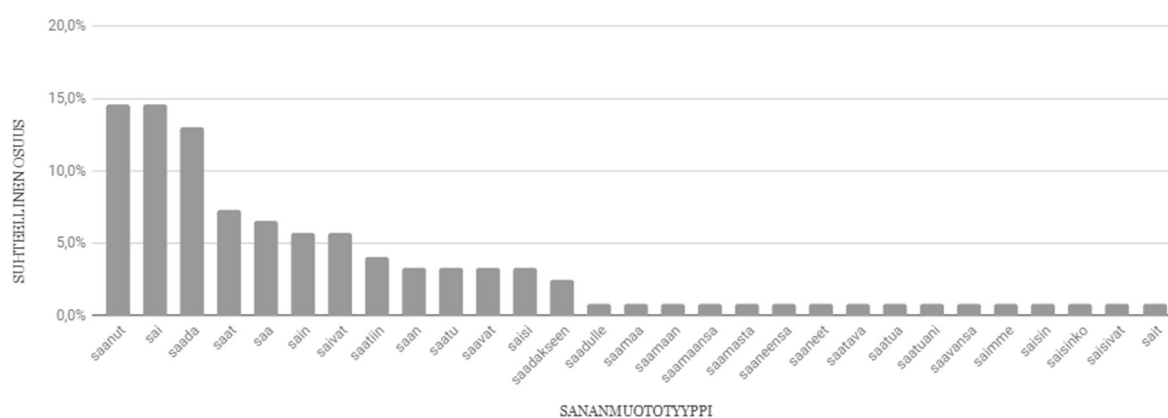
[157] Rahaa näihin palveluihin yhteiskunta **saa** veroista, jotka kansalaiset maksavat. (O)

[158] Hän oli kuullut että Norjassa voisi **saada** työn kalan perkaajana. (O)

[159] Asiakkaalla on myös oikeus **saada** palvelua viikonloppuna. (O)

³¹ z=1,85, p>0,05.

Natiivisuomen aineistossa SAADA-verbillä on 29 sananmuototyyppiä (123 SAADA-verbini esiintymää), jotka on esitelty taajuusjärjestyksessä kuviossa 8 (ks. tarkempi taulukko liite 6). Muodot *saanut* (14,6 %, n = 18) ja *sai* (14,6 %, n = 18) ovat taajimmat natiivisuomen aineiston sananmuodot merkitysluokassa 1. Muut natiiviaineistossa useammin kuin kerran esiintyneet sananmuodot ovat *saada* (13,0 %, n = 16), *saat* (7,3 %, n = 9), *saa* (6,5 %, n = 8), *sain* (5,7 %, n = 7), *saivat* (5,7 %, n = 7), *saatiin* (4,1 %, n = 5), *saan* (3,3 %, n = 4), *saatu* (3,3 %, n = 4), *saavat* (3,3 %, n = 4), *saisi* (3,3 %, n = 4) ja *saadakseen* (2,4 %, n = 3). Loput sananmuodot esiintyivät vain kerran (0,8 % per sananmuototyyppi): *saadulle*, *saamaa*, *saamaan*, *saamaansa*, *saamasta*, *saaneensa*, *saaneet*, *saatava*, *saatua*, *saatuani*, *saavansa*, *saimme*, *saisin*, *saisinko*, *saisivat* ja *sait*.



Kuvio 8. SAADA-verbini sananmuodot (n = 123) natiivisuomen aineistosta merkitysluokasta 1.

Tarkemmin tarkasteltuna natiiviaineistossa esiintyvät sananmuototyypit (n = 29) jakautuvat niin, että 48,3 % (n = 14) on finiittisiä (*saan*, *saat*, *saa*, *saavat*, *sain*, *sai*, *saimme*, *saivat*, *saatiin*, *saisin*, *saisinko*, *saisi*, *saisivat*, *sait*) ja vastaavasti 51,7 % (n = 15) infiniittisiä (*saada*, *saadakseen*, *saamaa*, *saamaan*, *saamaansa*, *saamasta*, *saanut*, *saaneet*, *saaneensa*, *saatu*, *saatua*, *saatuani*, *saadulle*, *saatava*, *saavansa*). Lähes kolmasosa kaikista (31,0 %, n = 9) sananmuototyypeistä viittaa kolmanteen persoonaan (finiittisesti yksikön ja monikon aktiivissa: *saa*, *saavat*, *sai*, *saivat*, *saisi*, *saisivat*) ja infiniittisesti possessiivisuffiksilla: *saadakseen*, *saamaansa*, *saaneensa*), 20,7 % (n = 6) ensimmäiseen persoonaan (finiittisesti yksikön ja monikon aktiivissa: *saan*, *sain*, *saimme*, *saisin*, *saisinko*) ja infiniittisesti possessiivisuffiksilla: *saatuani*) ja 6,9 % (n = 2) toiseen persoonaan (finiittisesti yksikön aktiivissa: *saat*, *sait*). Finiittisistä sananmuototyypeistä (n = 14) 57,1 % on preesensissä (n = 8: *saan*, *saat*, *saa*, *saavat*, *saisin*, *saisinko*, *saisi*, *saisivat*) ja imperfektissä 42,9 % (n = 6: *sain*, *sait*, *sai*, *saimme*, *saivat*, *saatiin*). Preesensin suurempi osuus selittynee natiivisuomessakin sillä, että se pitää sisällään useamman moduksen

esiintymiä kuin imperfekti. Indikatiivin useampi aikamuoto vaikuttanee myös natiivisuomessa siihen, että finiittisistä sananmuototyypeistä indikatiivissa on 71,4 % (n = 10) ja konditionaalissa 28,6 % (n = 4). Myöskään natiivisuomessa tempuksista ei tarkastella perfektiä ja pluskvamperfektiä ja moduksista imperatiivia ja potentiaalia ei esiinny. Infinitiittiset sananmuototyypit jakautuvat A-infinitiiviin (I infiniitti: *saada, saadakseen*), MA-infinitiiviin (III-infinitiivi: *saamaa, saamaan, saamaansa, saamasta*), NUT-partisiippiin (II partisiippi: *saanut, saaneet, saaneensa*) ja VA-partisiipin aktiiviin (I partisiippi: *saavansa*) ja passiiviin (IV partisiippi: *saatava*).

Aineiston perusteella voi siis sanoa, että natiivisuomen morfologinen priming SAADA-verbille merkitysluokassa 1 on *saanut* (natiivisuomen aineistoesimerkit 160–161) ja *sai* (natiivisuomen aineistoesimerkit 162–163). Natiivisuomelle näyttäisi olevan todennäköisempää, että verbin sananmuototyyppi on infiniittinen kuin finiittinen. Ottaen huomioon, että taajimmat sananmuototyypit (eli sanetasolla ero oppijansuomeen on vieläkin suurempi) viittaavat menneeseen aikamuotoon, voi natiivisuomessa SAADA-verbin nähdä keskittyvän jossain määrin menneisyyteen, vaikka sananmuototyyppien tasolla tilastollinen ero ei ollutkaan merkitsevä.

[160] Jo vuonna 1982 oli tullilaboratorio **saanut** oman ATK-järjestelmänsä tutkittujen tuotteiden ja tutkimustulosten rekisteröimiseksi. (N)

[161] [– –] eikä Koskinen ollut **saanut** koko aikana suun vuoroa. (N)

[162] Seuraavana päivänä Esku **sai** postia.

[163] Käytettäessä taas sellaista ruotsalaista laivaa, joka oli vähintään 50 lästin suuruinen, **sai** muustakin kaupasta täysvapauden. (N)

Oppijansuomessa finiittisiä sananmuototyyppisiä on enemmän kuin infiniittisiä ja natiivisuomessa infiniittisiä muototyyppisiä on enemmän kuin finiittisiä. Finiittisten tai infiniittisten³² osuuksien erot eivät ole kuitenkaan tilastollisesti merkitseviä. Natiivisuomessa infiniittisistä muototyypeistä esiintyvät myös VA-partisiipit (I ja IV partisiipit), joita ei ole oppijansuomen aineistossa. Toisin kuin oppijansuomessa, natiivisuomessa sananmuototyypeissä näkyy myös yksikön toinen persoona, mutta toisaalta puuttuu passiivin ja monikon ensimmäisen persoonan preesens, jotka esiintyvät oppijansuomessa. Monikon toista persoonaa ei esiinny kummassakaan aineistossa. Natiivisuomessa ero preesensien ja imperfektien ei ole aivan yhtä suuri kuin oppijansuomessa, mutta ero on hyvin pieni: ei tilastollisesti merkitsevä³³. Finiittisten muototyyppien jakautuminen indikatiiviin ja konditionaaliin oppijansuomen ja natiivisuomen välillä ei myöskään eroa: ero on tilastollisesti ei-merkitsevä³⁴.

³² z=0,95, p>0,05.

³³ z=-0,23, p>0,05.

³⁴ z=-0,33, p>0,05.

Seuraavaksi siirrytään sananmuotojen laadullisesta tarkastelusta sananmuotojen taajuuksien tilastollisiin eroihin. Kuviossa 9 on kuvattu oppijansuomen ja natiivisuomen aineistoissa esiintyvien sananmuototyyppien taajuuserot tilastollisesti merkitsevimmästä vähiten merkitsevään ylhäältä alas. Liitteessä 7 on tarkempi taulukointi z-arvoineen ja keskivirheineen.

Käyttöero sananmuodolla *saat*³⁵ on tilastollisesti merkitsevin, ja se onkin sananmuotojen ainoa erittäin merkitsevä tilastollinen eroavaisuus (natiivisuomen aineistoesimerkit 164–166). Tämä selittyy sillä, ettei oppijansuomessa esiinny sananmuotoa *saat* lainkaan.

[164] **Saat** tämän kasetin itsellesi, Topi sanoi. (N)

[165] Ei ole siis tarkoitus sanoa, että sinä tarvitset kolmen kuukauden virkavapauden, vaan sinä **saat**. (N)

[166] – **Saat** sitten illalla jotain terveellisempää, vaikka määmiä. (N)

Sananmuodot *saa*³⁶ (yliedustuminen oppijansuomessa), *saanut*³⁷ (aliedustuminen oppijansuomessa) ja *saavat*³⁸ (yliedustuminen oppijansuomessa) eroavat tilastollisesti merkittävästi. Melkein merkitseviä ovat sananmuodot *saivat*³⁹ (aliedustuminen oppijansuomessa), *saatu*⁴⁰ (aliedustuminen oppijansuomessa) ja *saamme*⁴¹ (yliedustuminen oppijansuomessa). Loppujen sananmuotojen käyttöerot aineistojen välillä eivät ole tilastollisesti lainkaan tai läheskään merkitseviä⁴², eli voidaan päätellä, että näiden sananmuotojen käyttö ei juurikaan eroa oppijan- ja natiivisuomen välillä.

Aineistoissa esiintyy yhteensä 33 sananmuototyyppiä. Näistä natiivisuomessa esiintyy 29 ja oppijansuomessa 21. Molemmille yhteisiä muototyyppisiä ovat *saa*, *saanut*, *saavat*, *saivat*, *saatiin*, *saada*, *saaneet*, *sai*, *saimme*, *saisin*, *saisivat*, *sain*, *saan*, *saisi*, *saadakseen*, *saamaan* ja *saamaansa*. Muodot, jotka esiintyvät oppijansuomessa, mutta eivät natiivisuomessa, ovat *saamme*, *saadaan*, *saaman* ja *saamastani* ja muodot, jotka esiintyvät natiivisuomessa, mutta eivät oppijansuomessa, ovat *saat*, *saatu*, *saadulle*, *saamaa*, *saamasta*, *saaneensa*, *saatava*, *saatua*, *saatuani*, *saavansa*, *saisinko* ja *sait*.

Huomattavaa ja kenties oletusarvoistakin on, että natiivisuomessa on enemmän sananmuototyyppisiä (n = 29) kuin oppijansuomessa (n = 21), vaikka merkitysluokasta 1 koottu natiivisuomen aineisto on pienempi (oppijansuomen aineisto n = 144, natiivisuomen aineisto n =

³⁵ z=-3,30, p<0,001.

³⁶ z=3,09, p<0,01.

³⁷ z=-2,98, p<0,01.

³⁸ z=2,89, p<0,01.

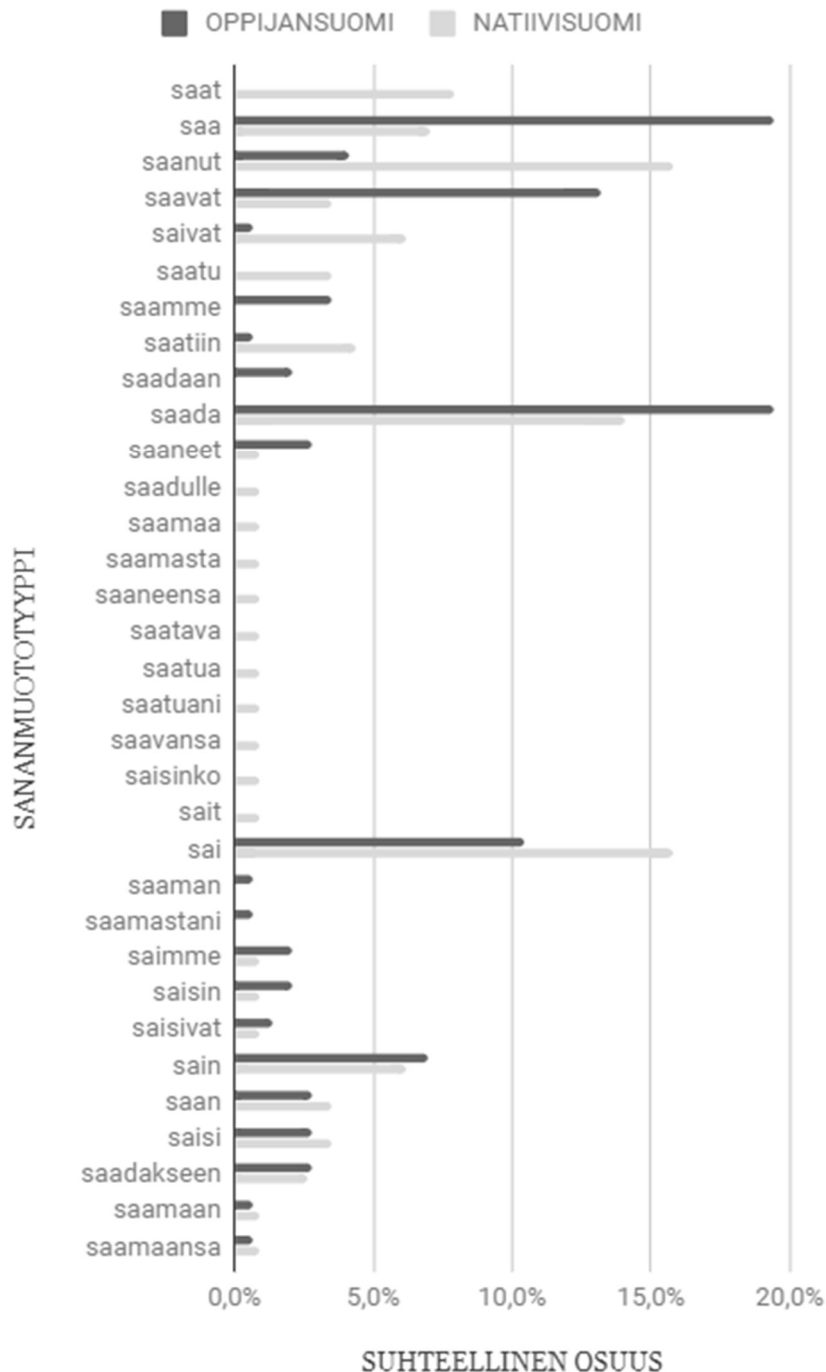
³⁹ z=-2,39, p<0,05.

⁴⁰ z=-2,18, p<0,05.

⁴¹ z=2,09, p<0,05.

⁴² z=-1,96–1,96, p>0,05.

123). Tilastollisesti ero jää piirun verran siitä, että se olisi melkein merkitsevä⁴³, mutta sitä voinee pitää tilastollisesti suuntaa antavana. Tästä voi päätellä, että natiivisuomessa SAADA-verbillä on hieman suurempi morfologinen variaatio merkitysluokassa 1 kuin oppijansuomessa samassa merkityksessä.



Kuvio 9 Oppijansuomen (144 sananmuotoa) ja natiivisuomen (123 sananmuotoa) SAADA-verbien sananmuotojen vertailu järjestettynä tilastollisen merkitsevyyden⁴⁴ mukaan.

⁴³ $z=-1,88$, $p>0,05$.

⁴⁴ Raja-arvot: 1,96 melkein merkitsevä ($p<0,05$); 2,57 merkitsevä ($p<0,01$); 3,29 erittäin merkitsevä ($p<0,001$).

6.3 SAADA-verbin semanttinen priming

Semanttisen primingin aineisto on koottu morfologisen primingin aineistosta siten, että sellaiset sananmuodot, jotka esiintyvät aineistossa vain yhdestä kolmeen kertaan on jätetty pois. Ennen tätä ortografiset virheet on korjattu (oppijankielen aineistossa *soi* → *sai*, *saame* → *saamme*, *saanut* → *saanut* ja *saanet*, *saanneet* → *saaneet*). Saman merkityksen vuoksi oppijankielen aineistossa sananmuoto *saadaksensa* on yhdistetty sananmuotoon *saadakseen* ja natiivikielen aineistossa *sais* on laskettu mukaan sananmuotoon *saisi*. Näin ollen oppijankielen 144 virkkeen aineistosta on poistettu 17 virkettä (sananmuodot *saadaan* x3, *saamaan*, *saamaansa*, *saaman*, *saamastani*, *saatiin*, *saimme* x3, *saisin* x3, *saisivat* x2 ja *saivat*) ja natiivikielen aineistosta 19 virkettä (sananmuodot *saadakseen* x3, *saadulle*, *saamaa*, *saamaan*, *saamaansa*, *saamasta*, *saaneensa*, *saaneet*, *saatava*, *saatua*, *saatuani*, *saavansa*, *saimme*, *saisin*, *saisinko*, *saisivat* ja *sait*). Jäljelle jää oppijansuomen aineistoon 475 ja natiivisuomen aineistoon 373 semanttisesti analysoitavaa sanaa tai ilmausta. Kaikki sanat kummastakin aineistosta on jaoteltu semanttisiin luokkiin liitteessä 8.

6.3.1 Semanttisten luokittelun periaatteet

Semanttisen analyysin aineisto on muodostettu morfologisen primingin aineistosta. Morfologisen primingin analyysissa mukana olevien virkkeiden SAADA-verbin sananmuodon ympäriltä on luokiteltu kaksi sanetta molemmin puolin, mutta vain saman virkkeen sisältä. Oppijansuomen semanttisen analyysin aineistossa on näin 475 sanetta ja natiivisuomen aineistossa 373. Luokittelun tukena on käytetty sanakirjoja (kuten KS ja TT), VISK:a ja Pajusen (2001) verbi-luokittelua.

Semanttiset luokat on muodostettu siten, että jokainen sana on luokiteltu omassa kontekstissaan ja sanaluokkarajojen yli. Pyrkimyksenä on ollut, että luokat nousisivat aineistoista eikä mitkään ennakko-oletukset tai muut rajoitukset ohjaisi luokituksen syntymistä. Tämän vuoksi useammin esiintyvä sama sana saattaa olla eri virkkeessä luokissa: polyseemisyys on siis otettu huomioon semanttisessa luokittelussa, mikä on mahdollista vain kontekstin huomioon ottaen. Esimerkiksi *sitten* on luokiteltu oppijansuomessa viidesti luokkaan 'aika, ikä' (oppijansuomen aineistoesimerkki 167) ja kerran luokkaan 'kausaalisuus' (oppijansuomen aineistoesimerkki 168). Lisäksi se esiintyy kerran natiiviaineistossa sellaisessa merkityksessä, jota ei ole voinut luokitella (natiivisuomen aineistoesimerkki 169), koska se on virkkeessä painottomana

täytesanana (KS s.v. *sitten*). Jokainen analysoitava sane on kuitenkin aina vain yhdessä semanttisessa luokassa.

[167] **Sitten** hän saa se summa, mutta krrrot (*korot) vähennetään jo ennen. (O)

[168] Mutta jos olisi laki, joka sanoo jos et halua mennä töihin ja et tee mitään **sitten** et saa rahaa valtiosta, laiskat menisivät varmasti työhön. (O)

[169] Vuonna 1852 Suomi **sitten** sai uuden tuontitullitaksan [– –]. (N)

Luokittelussa on noudatettu Pirkolan (2016) tapaan kapea-alaisuuden periaatetta, eli sana on ensisijaisesti luokiteltu kapea-alaisempaan, merkitykseltään tarkimpaan luokkaan, jos sopivia luokkia on ollut useampia. Koska analyysi koostuu kahdesta eri aineistosta ja luokat muodostuvat kummassakin aineistolähtöisesti, semanttisissa luokissa on aineistojen välillä hieman eroja. Jos toisessa aineistossa on muodostunut luokka, jota toisessa aineistossa ei ole muodostunut, tämän luokan sanat on sijoitettu laaja-alaisempaan luokkaan. Esimerkiksi oppijansuomen aineistoon muodostui luokka 'oikeus, mahdollisuus', johon on luokiteltu sana *lupa*. Sama sana on natiivisuomen aineistossa luokiteltu luokkaan 'abstrakti entiteetti', koska luokkaan 'oikeus, mahdollisuus' ei tullut tarpeeksi sanoja.

Fraseologisten periaatteiden mukaisesti useamman sanan muodostama selkeä kokonaisuus on luokiteltu yhtenä yksikkönä. Tällaisia ovat esimerkiksi etu- ja sukunimi -yhdistelmät (esimerkiksi *Vladimir Putin*) sekä muut kiinteät kokonaisuudet (esimerkiksi *tänä päivänä*). Samankaltainen luokitteluun liittyvä seikka on, että OLLA-verbin luokitteluun liittyy muutamia linjauksia. Ensinnäkin, mikäli OLLA-verbi on apuverbinä, on se merkitty köyhän merkityssällön vuoksi 'luokittelemattomaksi'. Sen sijaan jos OLLA-verbi ilmaisee olemista tai omistamista, on se luokassa 'tila'. Poikkeuksena OLLA-verbi on luokiteltu apuverbinä luokkaan 'ihmiset', mikäli se on ollut ainoa kohta virkkeessä, jossa näkyy tekijä (oppijansuomessa yksi tapaus *olen* ja natiivisuomessa yksi tapaus *olin*).

Semanttisten luokkien koot on rajattu niin, että yhteen luokkaan täytyy olla esiintymiä 1 % verran semanttisesti luokitelluista sanoista. Tämä tarkoittaa sitä, että oppijansuomen aineistossa minimimäärä on viisi esiintymää ja natiivisuomen aineistossa neljä. Tämän vuoksi, vaikka aineistoissa oli terveyteen liittyviä sanoja, ei ole kuitenkaan omaa luokkaa 'terveys' – esiintymät eivät ylittäneet yhden prosentin rajaa. Jotkut luokat ovat silti melko pieniä, mutta tiukemalla rajauksella luokittelemattomien ja laajojen luokkien koot olisivat kasvaneet suhteettomasti. Tällä rajauksella oppijansuomeen muodostui 24 semanttista luokkaa sekä joukko luokittelemattomia sanoja. Natiiviaineistosta oli havaittavissa 21 semanttista luokkaa ja joukko luokittelemattomia sanoja. Natiivisuomen aineiston kaikki semanttiset luokat löytyvät myös oppijansuomen aineistosta, mutta oppijansuomen aineistossa on kolme natiivisuomessa

esiintymätöntä luokkaa. Lisäksi molemmissa aineistoissa on sanoja, jotka eivät mene mihinkään määriteltyyn semanttiseen luokkaan, eivätkä muodosta keskenään omaa luokkaansa. Näitä sanoja käsitellään luokkana 'luokittelemattomat'. Molemmissa aineistoissa esiintyvät semanttiset luokat ovat 'ihmiset', 'aika, ikä', 'määrä, intensiteetti', 'konkreettinen entiteetti', 'relaationaalisuus', 'additiivisuus', 'paikka, suunta', 'kommunikaatio', 'modaalisuus', 'yhteiskunnallisuus, institutionaalisuus', 'negaatio', 'tila', 'raha', 'laatu, ominaisuus', 'prosessi', 'mentaalisuus', 'abstrakti entiteetti', 'ruoka', 'perhe', 'työ' sekä 'lahja, palkinto'. Lisäksi vain oppijansuomesta löytyvät luokat 'kausaalisuus', 'koulutus' ja 'oikeus, mahdollisuus'.

Luokkaan 'ihmiset' on luokiteltu henkilöiden nimet (esim. oppijansuomen aineistosta *Anssille, Kaisan, Minna Canthin* ja *Vladimir Putin*; natiivisuomen aineistosta *Asko, Petteri, Hannibal* ja *Varjokuu*) persoonapronominit (esim. oppijansuomen aineistosta *hän, he, sinulta* ja *minä* ja natiivisuomen aineistosta *meiltä, minun, häneltä* ja *sinä*), tekijäsanat (esim. oppijansuomen aineistosta *maahanmuuttajan* ja *osallistujat* ja natiivisuomen aineistosta *hoitaja, matkailija* ja *takavarikoijat*) ja muut ihmiseen viittaavat sanat (esim. oppijansuomen aineistosta *jokainen, ihmisiä, miehet, kukin, sankarit* ja *suomalaiset* ja natiivisuomen aineistosta *ei kukaan, henkilöitä, rengin* ja *kumमितä*). Kuten edellä kävi ilmi, luokassa on lisäksi kaksi OLLA-verbin esiintymää, kun lauseen subjekti on ollut näkyvässä vain apuverbissä. Luokassa 'ihmiset' oli muutamia rajatapauksia. Ensiksi edellä mainittu natiivisuomen aineiston *kumमितä*, joka tuntui menevän loogisesti luokkaan 'perhe'. Käyttöyhteys ei kuitenkaan viittaa perheeseen (natiivisuomen aineistoesimerkki 170). Vaikka esimerkissä puhutaankin 'pizzaperheestä', tulkitsemisen tapauksen pikemminkin kuvainnolliseksi kuin varsinaisesti 'perheeseen' liittyväksi, minkä vuoksi *kumमितä* on luokiteltu luokkaan 'ihmiset' eikä luokkaan 'perhe'. Toinen rajatapaus on *aikuiset*, joka voisi mennä myös luokkaan 'aika, ikä'. Natiivisuomen esimerkissä 171 käy kuitenkin ilmi, että kyse on pikemminkin ihmisryhmästä kuin varsinaisesti iästä (vrt. jos *aikuiset* esiintyisi lauseessa *Kasvoin aikuiseksi*, jolloin kyse olisi selkeämmin iästä).

[170] Pizzaperheen **kumमितä** don Salvatore sai sanoman Bronxin-päämajaansa, jossa hän edelleen oli sotilaineen patjoilla, että hänen ainoa tyttärensä Angelina ja hänen tuore vävynsä Jaripetteri olivat kadonneet haaksirikossa Etelämerellä. (N)

[171] Myös muut mahdollisesti luolalla olevat **aikuiset** saavat osansa. (N)

Luokassa 'aika, ikä' on nimensä mukaisesti aikaan ja ikään viittaavia ilmauksia, kuten vuosilukuja tai muuten tiettyyn hetkeen tai ikään, järjestykseen ja ehtimiseen liittyviä sanoja (esim. oppijansuomen aineistosta *syntymäpäivänä, uuden, tavallisesti, kuukausi, yhtäkkiä, joskus, 1970-luvulla, ennen, jälkeen* ja *aina* sekä natiivisuomen aineistosta *tänä päivänä, usein, kohta, vuoden, nyt, päätteeksi, sitten, illalla, ehdin, mennessä, kun* ja *jo*). Lisäksi luokassa 'aika,

ikä' on luokan 'määrä, intensiteetti' kanssa risteäviä sanoja, kuten *uudelleen, taas, kerralla* ja *joka kerta*, joissa on mukana sekä määrällinen ja ajallinen ulottuvuus. Tässä tutkimuksessa *uudelleen* ja *taas* on luokiteltu ajan (tarkemmin toistuvuuden) adverbeiksi ISK:n (§ 651) mukaan ja *kerralla* ja *joka kerta* ajanilmauksiksi Kielitoimiston ohjepankin (2015 § 663) mukaisesti. Muita rajatapauksia luokassa 'aika, ikä' on lekseemien UUSI ja VANHA sananmuodot (aineistoissa esim. *uuden, vanhan* ja *uutta*), jotka voisivat mennä myös luokkaan 'laatu'. Ne on kuitenkin sijoitettu luokkaan 'aika, ikä', koska 'uutuus' ja 'vanhuus' ovat vahvemmin 'aikaan' sidoksissa olevia käsitteitä kuin 'laatuun'. Merkitykseen 'laatu' tuntuu liittyvän enemmän lähi-merkitykset 'ominaisuus' ja 'arvo', kun taas UUSI ja VANHA eivät näitä välttämättä ole; sen sijaan ne liittyvät aina jotenkin aikaan. *Kun* voi merkitä 'aikaa' tai 'kausaalisuutta' (natiivisuomen aineistoesimerkki 172 vrt. oppijansuomen aineistoesimerkki 173):

[172] Tahdon proteesin mahdollisimman pian, ja **kun** sen saan, aloitan uudelleen. (N)

[173] Olen onnellinen, **kun** saan kukkia lahjaksi. (O)

Luokassa 'määrä, intensiteetti' on numeraaleja, numeroita ja vahvistavia sanoja (esim. oppijansuomen aineistosta *osa, viisi, paljon, liikaa, tarpeeksi, molemmat, miljoonien, oikein* ja natiivisuomen aineistosta *hieman, ihan, kaiken, 30* ja *vain*). Kokoon liittyvät sanat (SUURI ja PIENI) voisivat olla myös 'laatua' (kuten edellä UUSI ja VANHA), mutta koska PIENI viittaa koon vähyyteen (KS s.v. *pieni*) ja tällä logiikalla sen antonyymi SUURI koon paljouteen, on ne ryhmitelty luokkaan 'määrä, intensiteetti'.

Luokka 'raha' on nimensä perusteella melko läpinäkyvä. Siihen sisältyy sanana RAHA ja siihen läheisesti viittaavat sanat, kuten tuloihin, verotukseen ja rahan liikkumiseen liittyvät sanat (esim. oppijansuomen aineistosta *bafögia*⁴⁵, *rahaa, summa, veroista, palkkaa, lapsilisän, laina, ostaakseen, tuen, opintolainan* ja *rahoitusta* ja natiivisuomen aineistosta *remonttirahat, apurahoja, rahaa, lisätuloina, tullitaksan, taalarin* ja *rahapalkkaa*). Esimerkiksi verotukseen ja tullitaksaan liittyvät sanat olisivat käyneet myös luokkaan 'yhteiskunnallisuus, institutionaalisuus', mutta kapea-alaisuuden periaatteella ne on luokiteltu merkitykseltään kapea-alaisempaan ryhmään 'raha'.

'Additiivisuus' viittaa siihen, että jotain on tai tulee jonnekin lisää, jotakin yhdistyy tai että jotain on yhteensä jonkin verran (esim. oppijansuomen aineistosta *myös, sekä, ja, lisäksi, jne., eli* ja *vielä* sekä natiivisuomen aineistosta *ja, muutakin, yhteiseksi, mukaansa, kanssa* ja

⁴⁵ *Bundesausbildungsförderungsgesetz*, Saksan liittovaltion rahoitus opiskelijoille (Bundesministerium für Bildung und Forschung 2018).

kaiken kaikkiaan). *Eli* on luokassa 'additiivisuus' sillä perusteella, että se yhdistää kaksi ilmausta (KS s.v. *eli*). Luokka on alisteinen merkitykselle 'määrä, intensiteetti', mutta kapealaisempanakin se täyttää itsenäisen luokan kriteerit.

Luokassa 'negaatio' yhdistyy kielteisyyks ja vastakkaisuus. Erityisesti 'negaatiossa' on EI-verbin esiintymiä sekä MUTTA-konjunktiota, mutta myös muita 'vastakkaisuutta' ilmaisevia sanoja ja ilmauksia (esim. oppijansuomen aineistosta *et, ei, mutta, kuitenkin, päinvastoin, ei ole, paitsi* ja *kuitenkin* sekä natiivisuomen aineistosta *ei...ole, mutta, vaan, joka tapauksessa ja en*). Mukana on siis kieltoverbin esiintymät kokonaisuudessaan, koska osat EI ja OLLA esiintyvät usein niin lähekkäin, että EI-sanana yhteydessä oleva OLLA olisi hyvin usein jäänyt luokittelemattomaksi.

'Negaatio' on jo jossain määrin kieliopillinen luokka, mutta 'relaationaalisuus' on sitä vielä enemmän. Alun perin tähän luokkaan kuuluvat sanat olivat 'luokittelemattomissa'. Luokittelemattomien sanojen määrä alkoi kuitenkin paisua niin suureksi, että syntyi luokka 'relaationaalisuus', jonka sisältämät sanat ovat suorassa viittaussuhteessa johonkin toiseen kontekstin sanaan. Tähän kuuluvat siis relatiivi- sekä demonstratiivipronominit (esim. oppijansuomen aineisto *se, jotka, jota, niitä, mitä, jossa ja tämän* sekä natiivisuomen aineistosta *jonka, siitä, jostakin, tämän, tästä, sen ja siihenkin*). Vaikka kieliopillisuus onkin luokassa yhdistävä tekijä, kyse on myös samankaltaisesta, relaationaalisesta, merkityksestä.

'Paikka, suunta' pitää nimensä mukaisesti sisällään spatiaalisia sanoja (esim. oppijansuomen aineistosta *Hovimäki, Suomessa, Helsingistä, mistä tahansa, Etelä-Afrikassa, kotona, päälle* ja *Yhdysvalloissa* sekä natiivisuomen aineistosta *ratsastusleirille, sahatteollisuuskaupunki, Turun, takaisin, eteensä, apteekkiin, Vaasan ja Kotkan*). Kyse on siis konkreettisista paikoista ja suunnista, vaikka mukaan mahtuu myös metaforinen ilmaus matkasta (*kovimman kautta*). Abstraktimpaa luokkaa 'tila' käsitellään myöhemmin.

'Lahja, palkinto' pitää sisällään nimensä mukaisesti lahjaan ja palkintoon verrattavissa olevia asioita (esim. oppijansuomen aineistosta *joululahjoja, bonuspisteitä, perintönä, kansallisten stipendin (*kansallistipendin), lahjaksi* ja *Nobelin kirjallisuuspalkinnon* sekä natiivisuomen *palkinto, kruunun ja perintönä*). *Kruunun* on lajiteltu metaforisena luokkaan 'lahja, palkinto' sen kontekstin perusteella (natiivisuomen aineistoesimerkki 174):

[174] Luterilaisessa kulttuurissa kärsinyt kristallisoit ja jokainen osaa ulkoa sarjan aforismeja, joissa kovimman kautta saa kirkkaimman **kruunun**. (N)

Kyse on hyvin kapea-alaisesta luokasta, joka on alisteinen luokalle 'konkreettinen entiteetti', mutta johon riittää kuitenkin tarpeeksi sanoja muodostaakseen oman merkitykseltään tarkemman luokkansa.

'Kommunikaatio' sisältää sanoja viestinnästä, viestintävälineistä, viestin formaatista, mediasta ja kielestä (esim. oppijansuomen aineistosta *esittävä, kirjeen, suomea, puhelinnumero, uutisen ja puhua* sekä natiivisuomen aineistosta *vastauksen, suunvuoron, kuvanvaihdoissa, sanoman, postia, murteella ja luettavaa*). 'Kommunikaatio' lienee jossain määrin alisteinen luokalle 'mentalisuus', mutta kuitenkin kapea-alaisempi. Toisaalta 'kommunikaatio' voi olla myös konkreettisia tekoja, mitä 'mentalisuus' ei pidä sisällään, mikä puoltaa 'kommunikaation' olemassaoloa itsenäisenä merkitysluokkana.

Luokassa 'laatu, ominaisuus' on nimensä mukaisesti laatuun ja ominaisuuteen liittyviä sanoja. Vaikka luokittelu on tehty sanaluokkarajat ylittävästi, on (luonnollisesti) luokassa 'laatu, ominaisuus' vain adjektiiveja ja adverbeja (esim. oppijansuomen aineistosta *parempaa, harvinaisiin, kiva, hyvä, vaikea* ja *omaperäisiä* sekä natiivisuomen aineistosta *kirkkaimman, yksilöllisen, kelvollisia, turvallisempi, matalasti* ja *kylmän*). *Matalasti*-sanon kohdalla kyse on äänen laadusta (natiivisuomen esimerkki 175), minkä vuoksi se ei ole esimerkiksi luokissa 'määrä, intensiteetti' tai 'paikka, suunta'.

[175] – Haluan keskustella tuon taikurin kanssa, Koskinen sanoi **matalasti** ja sai vastaukseksi kylmän hymähdyksen: – Niin haluaisi moni muukin. (N)

'Abstrakti entiteetti' -luokan sanoja yhdistää aineettomuus (vrt. 'konkreettinen entiteetti'). *Entiteetti* on filosofian käsite, jonka merkitys on 'olio', 'olemus' tai 'itsenäinen kokonaisuus' (KS s.v. *entiteetti*). Kyse on laaja-alaisesta luokasta, jossa sanoja yhdistää vain niiden abstrakti olemus. Abstraktit entiteetit ovat siis aineettomia asioita (esim. oppijansuomen aineistosta *tapahtumien, palveluihin, todisteita, asioita, nimensä, urheilun* ja *tehtävän* sekä natiivisuomen aineistoista *lupa, etunimensä, tilaisuutta* ja *ihotautien*). Rajat ovat osin häilyviä, sillä *palveluihin* tai *tehtävän* voisivat myös olla luokassa 'prosessi'. Niiden ei kuitenkaan nähdä viittaavan 'prosessiin' yhtä vahvasti kuin 'abstraktiin entiteettiin', koska luokka 'prosessi' sisältää vaatimuksen jonkin asian tekemisestä tai muuttumisesta. Myös *urheilun* on rajatapaus: *urheilut* on luokassa 'prosessi', mutta siinä kyse on selkeämmin aktiivisesta toiminnasta (natiivisuomen esimerkki aineistoesimerkki 176), kun taas *urheilun* on kontekstissaan pikemminkin abstrakti välinearvo (oppijansuomen aineistoesimerkki 177):

[176] Isä oli nuorena voimistellut ja **urheilut**, saanut ihan palkintojakin, ja äiti oli lapsena voittanut koulunsa juoksukilpailut. (N)

[177] Vaikka vanhempi veli- Riku Riihilahti oli lapsena lupaava jalkapalloilija (jopa teini-ikäisenä hän sai sen **urheilun** avulla stipendin Yhdysvaltoihin! (O)

Luokka 'työ' on hyvin läpinäkyvä ja kapea-alainen luokka: sanat liittyvät kiinteästi työntekoon, uraan, ammattiin ja muihin työelämän ilmiöihin (esim. oppijansuomen aineistosta *työtä, virkan* (**viran*), *työskentele*, *ammatin* ja *ylennystäkään* sekä natiivisuomen aineiston *ammatin, kokouksen, virkaa* ja *työstään*). 'Työ' nähdään luokkana alisteinen luokalle 'yhteiskunnallisuus, institutionaalisuus' ja jossain määrin kytköksissä luokkaan 'raha'.

'Mentaalisuus' perustuu Pajusen (2001: 296) primääri-B-verbeihin. Pajusen (mp.) mukaan primääri-B-verbit ovat "mentaaliverbejä, jotka koodaavat ns. episteemistä avaruutta eli referentiaaliin asiantiloihin kohdistuvaa tietoa, sen lähdettä ja varmuusastetta". Luokkaan 'mentaalisuus' kuuluu Pajusen (2001) luokitukset 'psykologinen tila ja prosessi' ja 'perkeptio' (aistihavainnot). Pajusen (2001: 339) luokka 'puheakti' on erotettu luokkaan 'kommunikaatio', joka (kuten edellä kävi ilmi) on jossain määrin 'mentaalisuus'-luokan alainen, mutta kapea-alaisuuden periaatteen mukaisesti kuitenkin oma semanttinen luokkansa. Myös varmuusasteen ilmaukset on erotettu omaan luokkaansa 'modaalisuus'. 'Mentaalisuus' sisältää siis esimerkiksi oppijansuomen aineistosta sanoja *vihainen, vaikea, haluaa, onnellinen, mielestäni* ja *ilahduttava* sekä natiivisuomen aineistosta sanoja *myöntävän, mainetta, turhaan, toivomuksen* ja *vai-valla*. Sana *vaikea* on jälleen hyvä esimerkki siitä, kuinka konteksti on vaikuttanut jaotteluun. Oppijansuomen aineistoesimerkissä 178 *vaikea* on luokassa 'mentaalinen', kun taas oppijansuomen aineistoesimerkissä 179 *vaikea* on määriteltä luokkaan 'laatu, ominaisuus':

[178] Sitten se on kovin **vaikea** kansalle saa uuden valtionjohtajan. (O)

[179] Nykyaikana Venäjällä on **vaikea** saada hyvän työn. (O)

Esimerkissä 178 uuden valtionjohtajan saaminen kansalle nähdään vaikeana mentaalisenä prosessina, kun taas esimerkissä 179 *vaikea* tuntuu määrittelevän työnhaun laatua ja ominaisuutta Venäjällä – toisaalta tulkinta voi olla arvovalinnoista kiinni, koska jos esimerkki 179 olisi luokassa 'mentaalisuus', tulkinta voisi olla se, että työnhaku on ennen kaikkea mentaalinen prosessi, joka on hakijan päästä kiinni. Objektivisin valinta esimerkille 179 on siis 'laatu, ominaisuus', jolloin se nähdään nimenomaan määrittämässä työnhaun luonnetta ottamatta kantaa työnhaun prosessiin.

Kuten edellä kävi ilmi, 'modaalisuus' on määriteltä omaksi luokakseen. Se perustuu sellaiseen semanttiseen alueeseen, jolla ilmaistaan asian todenmukaisuuteen ja toteutumismahdollisuuksiin liittyviä arvioita. Modaalisuudella ilmaistaan asiantilan (epä)varmuutta, välttämättömyyttä, (epä)todennäköisyyttä, mahdollisuutta, mahdottomuutta, pakollisuutta, luvallisuutta tai

(epä)toivottavuutta. (VISK § 1551.) Perinteisten modaalisten verbien, adjektiivien ja adverbien lisäksi modaaliseksi on laskettu myös konjunktio *jos*, koska se ilmaisee muun muassa ehtolauseita ja mahdollisuutta (KS s.v. *jos*). Käytännössä luokassa 'modaalisuus' on siis varmuuden astetta indikoivia ilmauksia (esim. oppijansuomen aineistosta *ehkä, pitäisi, voisin, täytyy, toivottavasti* ja *tietysti* sekä natiivisuomen aineistosta *tulisi, jos, voitu, kuulemma, saattoi, taisi* ja *varmasti*).

Päinvastoin kuin luokassa 'abstrakti entiteetti', luokassa 'konkreettinen entiteetti' on fyysisesti olemassa olevia ja aistein havaittavia asioita. Luokat 'ruoka' ja 'lahja, palkinto' ovat alisteisia 'konkreettisen entiteetin' luokalle. 'Konkreettisia entiteettejä' ovat siis esineiden ym. lisäksi muut elolliset asiat (esim. 'eläimet' ja 'kasvit') paitsi 'ihmiset', joka on oma luokkansa. Luokkaan kuuluu esimerkiksi oppijansuomen aineistosta sanat *kalan, tavaroita, sähköä, kukkia, koiranpennun* ja *avaimet* sekä natiivisuomen aineistosta sanat *fluoria, lorukepin, varastojen, paketin, hiireltä, kasetin, narunpätkä, poroja* ja *partioveneensä*). Hankalia tapauksia ovat esimerkiksi *varastojen*, joka voisi olla jossain merkityksessä myös 'paikka, suunta' tai 'tila', mutta kuten natiivisuomen aineistoesimerkistä 180 paljastuu, kyse on pikemminkin jostain konkreettisesta säilytyspaikasta, joka ei kuitenkaan ole tarpeeksi täsmällinen sijainniltaan ollakseen 'paikka, suunta':

[180] jos Hannibal olisikin saanut **varastojen** täydennystä... (N)

'Tila' merkitsee staattista olemista ja pysymistä, mutta myös omistamista (ks. VISK § 1501; 1502; Pirkola 2016: 74). Tässä tutkimuksessa konkreettinen 'paikka' on erotettu abstraktimmasta 'tilasta' (esim. oppijansuomen aineistosta *on, säilyä, tilaa, vapautta, olosuhteita* ja *paenneelta* sekä natiivisuomen aineistosta *kaadetun, voi, lomaa, oman* ja *elämistä*). Kerrattakoon tässä kohti, että OLLA-verbi merkityksissä 'olla olemassa', 'elää', 'oleskella' ja 'omistaa' on määritelty luokkaan 'tila', apuverbinä sen sijaan se on jätetty luokittelematta vähäisen merkityksen vuoksi. 'Tila' on myös vastakkainen luokka luokalle 'prosessi', joka ilmaisee tapahtumasarjaa ja kehityskulkua (KS. s.v. *prosessi*).

'Yhteiskunnallisuus, institutionaalisuus' -luokassa on nimensä mukaisesti yhteiskuntaan ja instituutioihin liittyviä sanoja (pois lukien 'raha', 'työ' ja 'koulutus', jotka ovat omia luokkiin), kuten valtioon, uskontoon tai kulttuuriin liittyviä sanoja. Yksikään edellä mainituista merkityksistä ei muodosta omaa luokkaansa, joten ne ovat sen vuoksi merkityksen 'yhteiskunnallisuus, institutionaalisuus' alla. Luokkaan kuuluu esimerkiksi oppijansuomen aineistosta sanoja *valtiosta, yhteiskunta, kansalle, valtionjohtajan, viranomaisille, pankilta* ja *pääministeri* sekä

natiivisuomen aineistosta sanoja *sosiaaliturvaa, Suomi, kansainvälisessä, armeija, kaupasta, ortodoksiseen, perinteestä, viisumin ja tullikamarit*. *Suomi* on tässä yhteydessä merkitykseltään enemmän 'valtio' (natiivisuomen aineistoesimerkki 181) kuin 'paikka' (oppijansuomen aineistoesimerkki 182), minkä vuoksi se on luokassa 'yhteiskunnallisuus, institutionaalisuus' luokan 'paikka, suunta' sijaan. Myös *pääministeri* on kahden luokan rajalla, sillä se voisi olla myös luokassa 'ihmiset' tai 'työ'. *Pääministeri* koetaan kuitenkin ensisijaisesti yhteiskunnalliseksi toimijaksi (henkilö tittelin takana vaihtuu), minkä vuoksi se on päätynyt yhteiskunnallisia sanoja sisältävään luokkaan.

- [181] Kun päätullijohtokunta oli aloittanut toimintansa ja kun **Suomi** oli saanut oman tullitaksan ja kaupapasäännön, oli suuriruhtinaskunnan tulliolojen alkujärjestelyt saatu vuoden 1812 kuluessa päätökseen. (N)
 [182] Lopussa minä toivon, että voin opiskella suomea **Suomessa** ja saada työtä siellä. (O)

'Kausaalisuus' sisältää syitä ja seurauksia: KS (s.v. *kausaalisuus*) määrittelee 'kausaalisuuden' 'syysuhteisuudeksi', 'syysuhteeksi' ja 'kausaliteetiksi'. 'Kausaalisuus' esiintyy oppijansuomessa esimerkiksi sanoin *sitten, siis, koska, kun, siksi*. Natiivisuomessa ei esiintynyt tarpeeksi kausaalisia sanoja, jotta niistä olisi muodostunut oma luokkansa. Myös luokat 'oikeus, mahdollisuus' ja 'koulutus' esiintyvät yksinomaan oppijansuomenaineistossa. Nimensä mukaisesti 'oikeus, mahdollisuus' sisältää oikeuteen ja mahdollisuuteen liittyviä sanoja: *pääsyä, mahdollisuuden, oikeus, mahdollista* ja *tilaisuus* ja 'koulutus' koulutukseen liittyviä sanoja: *opiskelija, lukion, yliopistossa* ja *opiskella*. *Opiskelija* voisi kuulua myös luokkaan 'ihmiset', *lukio* ja *yliopisto* luokkaan 'yhteiskunnallisuus, institutionaalisuus' ja *opiskella* luokkaan 'prosessi', mutta laaja-alaisuuden periaatteen mukaisesti ne on luokiteltu tarkempiin merkitysluokkiin, kun yhden prosentin esiintyvyys on täyttynyt.

'Ruoka'-luokka sisältää syömiseen ja juomiseen liittyviä asioita. Esimerkiksi oppijansuomen aineistossa esiintyy sanoja *suklaamunat, alkoholi, lauantaimakkaraa* ja *kahvikuppi*. Natiivisuomen aineistossa merkitysluokkaan 'ruoka' kuuluu esimerkiksi sanat *äidinmaidon korviketta, mahani, söi* ja *bouillabaisse-keittoa*. *Mahani* kuuluu luokkaan 'ruoka', koska kyse on maahan syömisestä täyteen (natiivisuomen esimerkki 183):

- [183] Uni tuli peitepompan alla vasta aamun jo valostuessa, kun mieleni retkeili vapaudessa ja haaveilin kaikenlaisista hyvistä ruuista, joita voisin kohta saada syödäkseni **mahani** täyteen. (N)

Luokasta 'ihmiset' on erotettavissa kapea-alaisempi luokka 'perhe', jossa on nimensä mukaisesti perheeseen ja laajemmin sukulaisuuteen liittyviä sanoja. Oppijansuomen aineistossa

perhesanastoa on esimerkiksi *poika, vanhemmatkin, perheensä* ja *vaari* sekä natiivisuomen aineistossa esimerkiksi *vanhemmilleen, äidin, isä ja lapset*.

Pajusen (2001: 12) mukaan 'teko', 'tapahtuma' ja 'muutos' ovat prosesseja. Pirkolan (2016: 74) tapaisesti myös tässä tutkimuksessa luokassa 'prosessi' on sekä verbejä että prosesseiksi tulkittavia nomineja. Luokka on vastakkainen luokalle 'tila'. Oppijansuomen aineistossa tällaisia sanoja on esimerkiksi *yrittävät, saavat* ja *alkoi*. Natiivisuomen aineistossa luokkaan 'prosessi' kuuluu esimerkiksi sanoja *toimittuaan, täydennystä, tulla, nouseva, remontista, urheillut* ja *aloitan*. Oppijansuomen aineistoesimerkistä 184 on alleviivattu *saavat*, joka kuuluu kotekstiin eikä ole noodin asemassa (toisin kuin alleviivattu *saavat*):

[184] Kaiket, jotka saavat Bafögia **saavat** myös lapsilisää. (O)

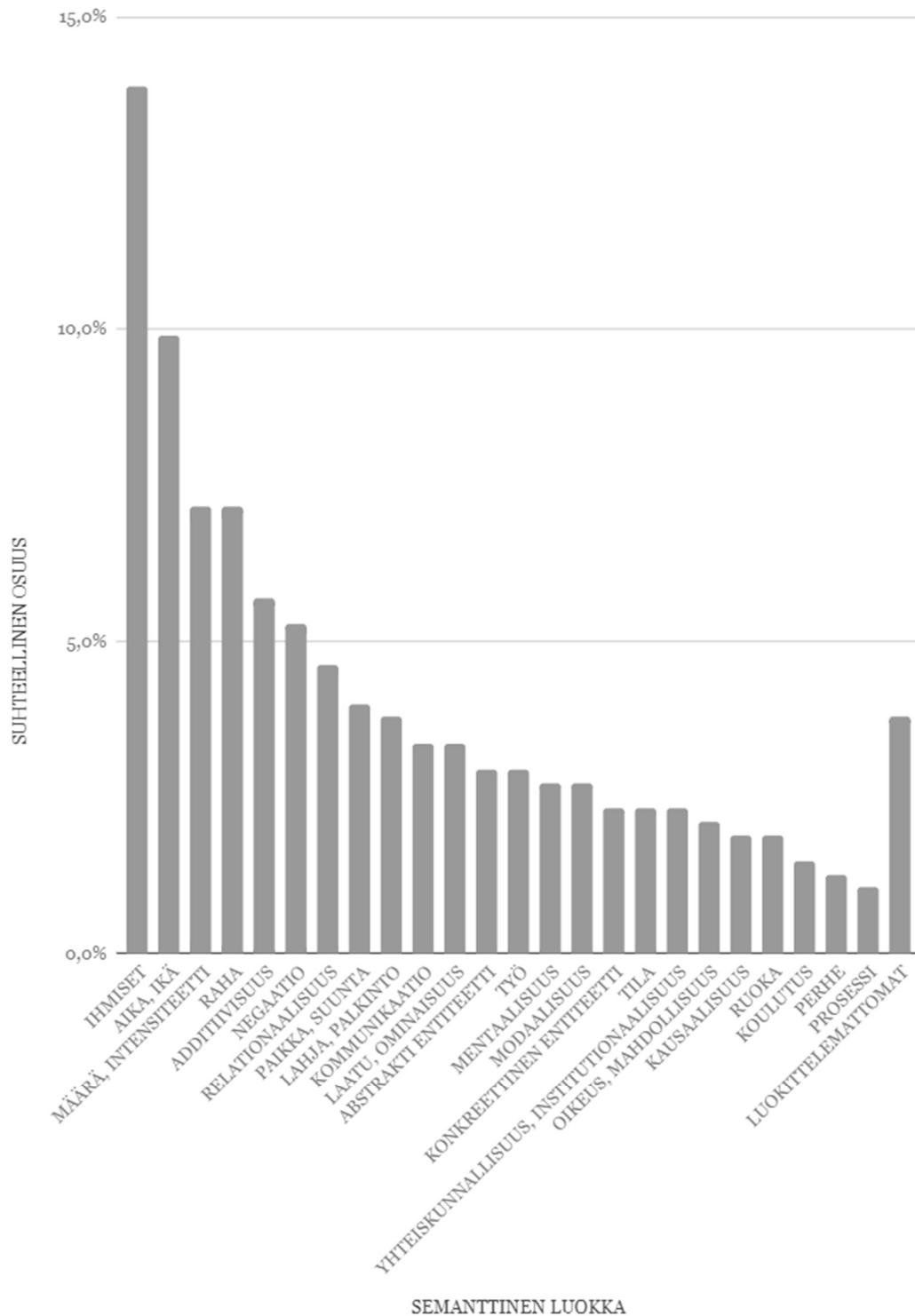
Viimeinen luokka on 'luokittelemattomat', jossa on sanoja, jotka eivät sopineet mihinkään edellä esiteltyyn luokkaan. Molemmissa luokissa luokittelematta on jäänyt konjunktio *että* ja OLLA-verbi, kun se on apuverbin asemassa (paitsi kaksi tapausta, joissa OLLA-verbi ainoana sanana ilmensi lauseen tekijää). Oppijansuomessa luokittelematta jäi myös *vaikkapa* ja natiivisuomessa *kuten, kuin, sitten* ja *takia*. *Takia* kävisi luokkaan 'kausaalisuus', jota ei kuitenkaan muodostunut natiivisuomen aineistoon, ja koska se ei sovi muihinkaan luokkiin, on se natiivisuomen aineistossa jäänyt 'luokittelemattomiin'.

6.3.2 SAADA-verbin semanttinen priming oppijan- ja natiivisuomessa

Tässä alaluvussa vastataan kolmanteen tutkimuskysymykseen. Oppijansuomessa semanttisia luokkia on 24 luokittelemattomat pois luettuna. Suuruusjärjestys taajimmasta alkaen on kuvattu kuviossa 10 (ks. tarkempi taulukko liite 9). 'Luokittelemattomat' on taajuudesta huolimatta viimeisenä hahmottamisen helpottamiseksi.

Oppijansuomen semanttisista luokista SAADA-verbin sisäisestä merkitysluokasta 1 (n = 475) suurin luokka on 'ihmiset' (13,9 %, n = 66), toiseksi suurin 'aika, ikä' (9,9 %, n = 47) ja jaetulla kolmannella sijalla ovat 'määrä, intensiteetti' (7,2 %, n = 34) ja 'raha' (7,2 %, n = 34). Loput luokat taajuusjärjestyksessä ovat 'additiivisuus' (5,7 %, n = 27), 'negaatio' (5,3 %, n = 25), 'relaationaalisuus' (4,6 %, n = 22), 'paikka, suunta' (4,0 %, n = 19), 'lahja, palkinto' (3,8 %, n = 18), 'kommunikaatio' (3,4 %, n = 16), 'laatu, ominaisuus' (3,4 %, n = 16), 'abstrakti entiteetti' (2,9 %, n = 14), 'työ' (2,9 %, n = 14), 'mentaalisuus' (2,7 %, n = 13), 'modaalisuus' (2,7 %, n = 13), 'konkreettinen entiteetti' (2,3 %, n = 11), 'tila' (2,3 %, n = 11), 'yhteiskunnallisuus,

institutionaalisuus' (2,3 %, n = 11), 'oikeus, mahdollisuus' (2,1 %, n = 10), 'kausaalisuus' (1,9 %, n = 9), 'ruoka' (1,9 %, n = 9), 'koulutus' (1,5 %, n = 7), 'perhe' (1,3 %, n = 6), 'prosessi' (1,1 %, n = 5) ja 'luokittelemattomat' (3,8 %, n = 18).



Kuvio 10. Oppijansuomen aineiston semanttiset luokat merkitysluokasta 1 (n = 475).

Kuten luokkien suuresta määrästä voi päätellä, ei oppijansuomen merkitysluokan 1 kontekstista nouse yhtä selkeää merkitysluokkaa, vaan monta pientä. Tarkempaan laadulliseen tarkasteluun otetaan kolme taajinta luokkaa (tai tässä tapauksessa neljä, koska kolmas sija on jätetty): 'ihmiset', 'aika, ikä', 'määrä, intensiteetti' ja 'raha'. Muitakin luokkia olisi syytä tarkastella tarkemmin, mutta tutkielman laajuuden huomioon ottaen se ei tässä tutkimuksessa ole mahdollista.

Koska kyse on verbistä, jonka valenssiin kuuluu kolme argumenttia, joista kaksi on henkilötarkotteisia (Pajunen 2001: 264), 'ihmiset' lienee luonnollinen semanttinen priming. Lisäksi luokan 'ihmiset' nousua suurimmaksi luokaksi selittää SAADA-verbin merkitys 1, jossa on kyse 'tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkun tai jonkin omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jokin tulee'. Tämän voi tulkita niin, että 'saaja' (henkilö) on läsnä jo verbin sisäisen merkityksen kuvauksessa. Edellä morfologisen primingin analysoinnissa näkyi, että oppijansuomen merkitysluokassa 1 SAADA-verbi suosii kolmatta persoonaa. Tämä näkyy myös semanttisten luokkien tasolla, sillä lähes kolmannes (28,8 %, n = 19) luokan 'ihmiset' (n = 66) sanoista on jokin lemmän HÄN sananmuodoista (oppijansuomen aineistoesimerkit 185–187).

[185] Nyt **hän** ei **saa ruokaa ennen**⁴⁶ kuin huomenna. (O)

[186] Lopuksi vuonna 1580 **hänen perheensä sai rahaa ja** vapauttiin Cervantes. (O)

[187] **Timo sai hänen kahvikuppi** ja sämpylä ja se maistui hyvältä. (O)

Jos HÄN-esiintymiin lisätään HE, kolmannen persoonan viittausten määrä luokassa 'ihmiset' (n = 66) on jo yli kolmasosa (36,4 %, n = 23). Lisäksi joukossa on lukuisia sanoja, joiden kanssa verbi kongruoisi (yksikössä tai monikossa) kolmannessa persoonassa: *ihminen, ihmiset, F. E. Sillanpää, jokainen, jotkut, kaikki (*kaikki), kukin, miehet, mies, naiset, osallistujat, sankarit, sinkku, suomalaiset, Timo, venäläiset ja Vladimir Putin*. Vaikka ensimmäinen persoona näkyy morfologisesti, luokassa 'ihmiset' se ei ole yhtä ilmeinen: minuuteen viittaa vain viisi sanetta, *minulta, minä x2, minähän ja olen*. Ensimmäisen persoonan heikko edustus kolmanteen nähden selittynee sillä, että ensimmäisessä persoonassa subjektin voi pudottaa pois (*minä saan = saan*), kun taas kolmannessa persoonassa ei (*hän saa ≠ saa*).

Persoonien lisäksi verbit taipuvat aikamuodoissa, mikä näkyy SAADA-verbin merkitysluokan 1 toiseksi suurimpana semanttisena luokkana kontekstissa. 'Aika, ikä' -luokasta ei ilmene yhtä selkeästi kuin 'ihmiset'-luokassa, millaisia sanoja on eniten. Tarkemmin

⁴⁶ Tarkastelualue alleviivattu. Esimerkkisane semanttisesta luokasta ja SAADA-verbin sananmuoto tummennettu.

tarkasteltuna saneet voi kuitenkin jakaa menneisyyteen tai aikaviittauksesta taakse päin viittaaaviin ilmauksiin (*1970-luvulla, ennen, kauan sitten ja vielä*), tulevaisuuteen, ikuisuuteen tai aikaviittauksesta eteenpäin viittaaviin ilmauksiin (*aina, jatkossa, jälkeen, lähitulevaisuudessa, pian ja lopulta*), kiinteisiin kerran vuodessa toistuviin ajankohtiin (*jouluna, syntymäpäivä ja syntymäpäiväksi*), epämääräisiin ja toistuvuutta ilmaiseviin aikaviittauksiin (*joskus, toisinaan, usein x2, tavallisesti, joka kerta, taas ja ketkeksi (*hetkeksi)*), nykyhetkeen tai äkillisyyteen viittaaviin (*nyt x2 ja yhtäkkiä*), tiettyyn hetkeen viittaaviin (*silloin x2, sitten x4 ja kun*), tiettyyn aika väliin viittaaviin (*siinä *siinä ajassa*) vuoden jaksottamiseen ja ajan pituuteen liittyviin aikaviittauksiin (*kuukausi, päivällä, päivänä x2, viikonloppuna, vuodelta, vuotta x2 ja tänään*), ikäviittauksiin (*teini-ikäisenä ja vanhin*) ja uutuutta ja vanhuutta ilmaiseviin sanoihin (*uuden*). Näyttäisi siltä, että eniten aikaviittaukset ovat joko suhteessa johonkin muuhun aikaan (viittaukset menneisyyteen, tulevaisuuteen, nykyisyyteen) (oppijansuomen aineistoesimerkit 188–189) tai ilmaisevat toistuvuutta (oppijansuomen aineistoesimerkit 190–191).

[188] Minähän sain aika kauan sitten tuo sähköposti sinulta, missä olit kirjoittanut siitä kurssista. (O)

[189] Olen aivan varma siitä että me kaikki kolme kurssia jotka opimme hetkellä suomea ylopistossa (*yliopistossa) saamme lähitulevaisuudessa helposti töitä. (O)

[190] Usein naiset saavat pienempää palkkaa kuin miehet. (O)

[191] Jos on mahdollisuus, haluan saada kansallis-stipendin (*kansallisstipendin) taas!

’Määrä, intensiteetti’ -luokassa on myös eroteltavissa alakategorioita. Luonnollisesti luokassa on runsaasti numeraaleja tai tarkasti lukumäärän määrittäviä sanoja (*kaksi x2, miljoonien, molemmat, viisi x2, yhden, yksitoista ja toisen x2*), epätarkasti määrään tai osuuteen viittaavia sanoja (*aika paljon, paljon x2, muutamien, pienen, osa, lasin ja prosenttia*), määrää vertailevia sanoja (*enemmän x4, pienempää, yhä enemmän ja suurin*), määrään kantaa ottavia sanoja (*liikaa x3 ja tarpeeksi*), sanoman voimakkuutta (intensiteettiä) lisääviä vahvistussanoja (*aika, oikein ja niin*) sekä määrää rajoittavia (*vain*) että määrän saavuttamiseen viittaavia sanoja (*kokoon (*kokoon)*). Tarkat (oppijansuomen aineistoesimerkit 192–194) ja epätarkat (oppijansuomen aineistoesimerkit 194–195) määrän ilmaukset näyttäisivät muodostavan suurimman osan luokan sanoista.

[192] Mutta kysyin ja pyysin lisää ja kun täytin yksitoista vuotta, sain lopulta kaksi marssu (*marsua). (O)

[193] Tänään sain viisi arpalippua, mutta ilman voittoa. (O)

[194] [– –] molemmat olivat saanet (*saaneet) kokoon (*kokoon) aika paljon rahaa. (O)

[195] Käännökset tehdään kotona ja saamme joka kerta pienen osan. (O)

Luokassa ’raha’ on runsas määrä rahaan ja sen määrään itseensä viittaavia sanoja (*rahaa x12, rahallisesti, rahallista ja summa*), yhteiskunnalta saataviin tukiin tai lainaan viittaavia

sanoja (*bafögia* x3, *opintolainan*, *opintorahaa*, *lapsilisän*, *tuen* ja *laina*), yhteiskunnalle maksettaviin asioihin viittaavia sanoja (*veroista* x2), rahan ansaitsemiseen liittyviä sanoja (*palkkaa* x3, *palkkiota* ja *palkkoja*), rahoitukseen liittyviä sanoja (*jatkorahoitushakemuksia* ja *rahoitusta*) ja yksittäiset valuuttaan ja sen käyttöön liittyvät sanat *kultaa* ja *ostaakseen*. Raha korostuu sanana, ja suurimpina alakategorioina voidaankin nähdä raha itse (oppijansuomen aineistoesimerkit 196–197), mutta myös yhteiskunnan tuet (oppijansuomen aineistoesimerkit 196; 198) ja rahan ansaitsemisen (oppijansuomen aineistoesimerkki 199). Luokan sanat ovat pääosin partitiivissa, mikä viittaisi niiden olevan SAADA-verbin kohteena, mikä toki on oletusarvoistakin (*saada rahaa*).

[196] Minun vanhempani eivät voisi auttaa minua rahallisesti, mutta saan opintolainan, jonka kanssa maksan minun vuokraani ja kaikki muut elinkustannukseni. (O)

[197] Venäläisiltä suomalaiset saavat rahaa, mitä on edullista suomalaiselle taloudelle. (O)

[198] Suurin osa saa Bafögia. (O)

[199] Kun opiskelee, oppii asioita ja saa parempaa palkkaa. (O)

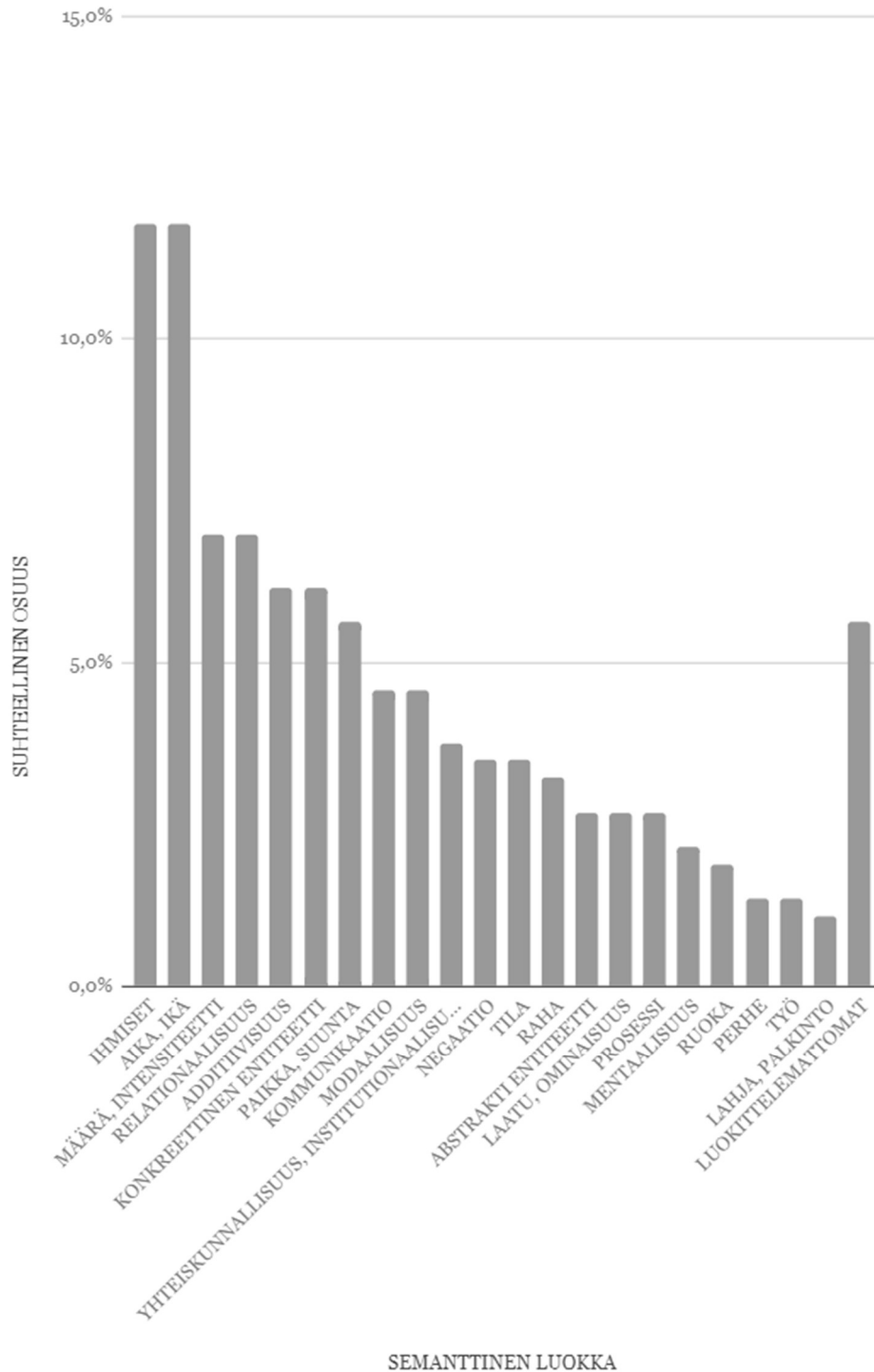
Natiivisuomen aineistossa semanttisia luokkia on 21 luokittelemattomat pois lukien. Merkitysluokan 1 semanttiset luokat SAADA-verbin kotekstista on kuvattu taajuusjärjestyksessä kuviossa 11 (ks. tarkempi taulukko liite 10). 'Luokittelemattomat' on myös tässä kuviossa taajuusjärjestyksestä huolimatta viimeisenä hahmottamisen helpottamiseksi.

Taajimmat luokat natiivisuomen aineistossa (n = 373) ovat 'ihmiset' (11,8 %, n = 44) ja 'aika, ikä' (11,8 %, n = 44), toiseksi taajimmat 'määrä, intensiteetti' (7,0 %, n = 26) ja 'relatiivisuus' (7,0 %, n = 26). Loput merkitysluokat taajuusjärjestyksessä ovat 'additiivisuus' (6,2 %, n = 23), 'konkreettinen entiteetti' (6,2 %, n = 23), 'paikka, suunta' (5,6 %, n = 21), 'kommunikaatio' (4,6 %, n = 17), 'modaalisuus' (4,6 %, n = 17), 'yhteiskunnallisuus, institutionaalisuus' (3,8 %, n = 14), 'negaatio' (3,5 %, n = 13), 'tila' (3,5 %, n = 13), 'raha' (3,2 %, n = 12), 'abstrakti entiteetti' (2,7 %, n = 10), 'laatu, ominaisuus' (2,7 %, n = 10), 'prosessi' (2,7 %, n = 10), 'mentaalisuus' (2,1 %, n = 8), 'ruoka' (1,9 %, n = 7), 'perhe' (1,3 %, n = 5), 'työ' (1,3 %, n = 5), 'lahja, palkinto' (1,1 %, n = 4) ja 'luokittelemattomat' (5,6 %, n = 21).

Tutkielman laajuuden puitteissa on mahdollista käsitellä vain kolmea taajinta luokkaa tarkemmin laadullisesti, mutta kuten oppijansuomessakin, on natiivisuomessa jaettuja sijoja. Tämän vuoksi käsittelyssä on neljä taajinta 1. ja 2. sijat jakavat 'ihmiset', 'aika, ikä', 'määrä, intensiteetti' ja 'relatiivisuus'.

Luokan 'ihmiset' sanoissa korostuu nimet (*Petteri*, *Asko*, *Hannibal*, *Esku*, *Hilarioiden*, *Varjokuu*, *Syysprinssi*, *Koskinen*, *Sluppi*, *don Salvatore*, *Lomakka* ja *Pauliina*). Nimet kattavat lähes kolmasosan (27,3 %, n = 12) luokan 'ihmiset' sanoista (n = 44). Muut viittaukset ihmisiin

koostuvat kolmannen persoonan pronomineista (*hän* x4, *häneltä* x2 ja *he* x2), ensimmäisen persoonan pronomineista (*minun*, *minä* x2, *minulta*, *mia*, *minäkin*, *meiltä* ja *olin*), toisen persoonan pronomineista (*sinä* x3), muista pronomineista (*jokainen*, *joku* ja *ei kukaan*) sekä suorista ihmisviittauksista, rooleista ja tekijöistä (*hoitaja*, *takavarikoijat*, *rengin*, *emännältä*, *aikuiset*, *uhrit*, *matkailija*, *kummisetä*, *henkilöitä* ja *ihmisessä*).



Kuvio 11. Natiivisuomen aineiston semanttiset luokat merkitysluokasta 1 (n = 373).

Vaikka nimet kattavat lähes kolmasosan luokan sanoista, yhdistämällä kaikki pronominit huomataan, että yhteensä ne kattavat puolet (50,0 %, n = 22) luokan 'ihmiset' sanoista (n = 44). Natiivisuomen aineistoesimerkit 200–202 havainnollistavat nimiviittauksia kontekstissa:

[200] Mutta Asko yritti saada rahaa vanhemmilleen, joilla oli vaikeuksia. (N)

[201] Kapakassakin Sluppi sai oman tuolinsa [– –]. (N)

[202] Juhla- ja syntymäpäivinä Hilarioiden lapset saivat vanhempien kanssa myös lasin samppanjaa. (N)

Luokassa 'aika, ikä' ei ole selkeitä alaluokkia. Jotakin luokittelua on kuitenkin mahdollista tehdä: luokassa on sanoja, jotka viittaavat aiempaan hetkeen (*ensin* x2, *eilän* (**eilen*) ja *juuri*), nykyhetkeen (*tänä päivänä*, *jo* ja *nyt* x2), tulevaisuuteen (*kohta*, *vastaisuudessakin* ja *seuraavana päivänä*), vuoden ja päivän jaksottamiseen ja ajan pituuteen (*vuoden*, *vuonna* x2 ja *illalla*), jonkin loppumiseen (*päätteeksi*, *enää*, *mennessä*, *kuluttua* ja *lopullisia*), tiettyyn ajanjaksoon (*aikana*, *aikaa*, *aikoihin* ja *respiittiaikaa*), vuosilukuun (*1558*, *1780* ja *1996*), ajalliseen toistuvuuteen (*taas*, *usein* ja *uudelleen* x2), uutuuteen ja vanhuuteen (*uutta*, *uuden* x2 ja *vanhan*), tiettyyn hetkeen (*sitten* ja *kun*), toistuvaan ajankohtaan (*jouluaattona* x2), ajallisiin ja määrällisiin kertoihin (*kerralla*, *samalla* ja *vuoroo*) sekä yksittäisiin järjestykseen (*ensimmäisenä*) ja ehtimiseen (*ehdin*). Koska yksittäisiä käyttötapoja ei ole havaittavissa, voidaan päätellä natiivisuomessa olevan semanttisen luokan 'aika, ikä' koostuvan laajasta variaatiosta.

'Määrä, intensiteetti' -luokassa (n = 26) korostuu pienuuteen ja vähyyteen liittyvät sanat (*vähän*, *vähemmän*, *pienempi*, *pienen* x2 ja *hieman*), jotka kattavat luokan sanoista lähes neljänneksen (23,1 %, n = 6). Loput luokan sanat liittyvät suuruuteen (*suurempi*, *suuria* ja *suuren*), kokoon (*kokoon*, *koko*, *kaiken* ja *suuruinen*), osaan (*osansa* ja *osuutensa*) tarkkaan määrään (2, 2,5, *miljoonaa*, *kaksi* ja 30), ryhmän kokoon tai mittaan (*lästin*, *parttion* ja *joukon*), sanoman vahvistamiseen (*ihan* ja *kyllä*) ja yksittäistapauksena määrän rajoittamiseen (*vain*). Natiivisuomessa luokka 'määrä, intensiteetti' näyttää siis ensisijaisesti viittaavaan kokoon: 'pienuus' (huom: vähyyteen liittyvät sanat vähennetty) (natiivisuomen aineistoesimerkki 203), 'suuruus' (natiivisuomen aineistoesimerkki 204) ja (myös ryhmän) 'koko' (natiivisuomen aineistoesimerkki 204) (yht. n = 13) kattavat natiiviaineiston luokasta 'määrä, intensiteetti' yhteensä 50,8 %.

[203] Painelin heti apteekkiin ja sain pienen rasian, jossa oli katajanmarjan kokoisia pillereitä. (N)

[204] [– –] näytösoikeudenkäynnit saivat suuren joukon innokkaita puolestapuhujia [– –]. (N)

Luokan 'relationaalisuus' (n = 26) sanat ovat jaettavissa relatiivipronomineihin (*joissa, joka x3, jokin, jolla, jonka x2, jossa, jostakin, jotka*) ja demonstratiivipronomineihin (*ne, näiden, se x2, sen x3, siihenkin, siitä x4, tämän ja tästä*). Mielenkiintoista on, että relatiivipronomineista esiintyy vain JOKA (natiivisuomen aineistoesimerkki 205). Relatiivipronominin katavat luokasta 46,2 % (n = 12) ja demonstratiivipronominin 53,8 % (n = 14) (natiivisuomen aineistoesimerkki 206). Vaikka demonstratiivipronomineja on enemmän, ero ei ole suuri. Luokan 'relationaalisuus' nousu natiivisuomen jaetulle toiselle sijalle kertonee siitä, että natiivisuomen aineistossa on kompleksisia ja paljon viittaussuhteita sisältäviä virkkeitä.

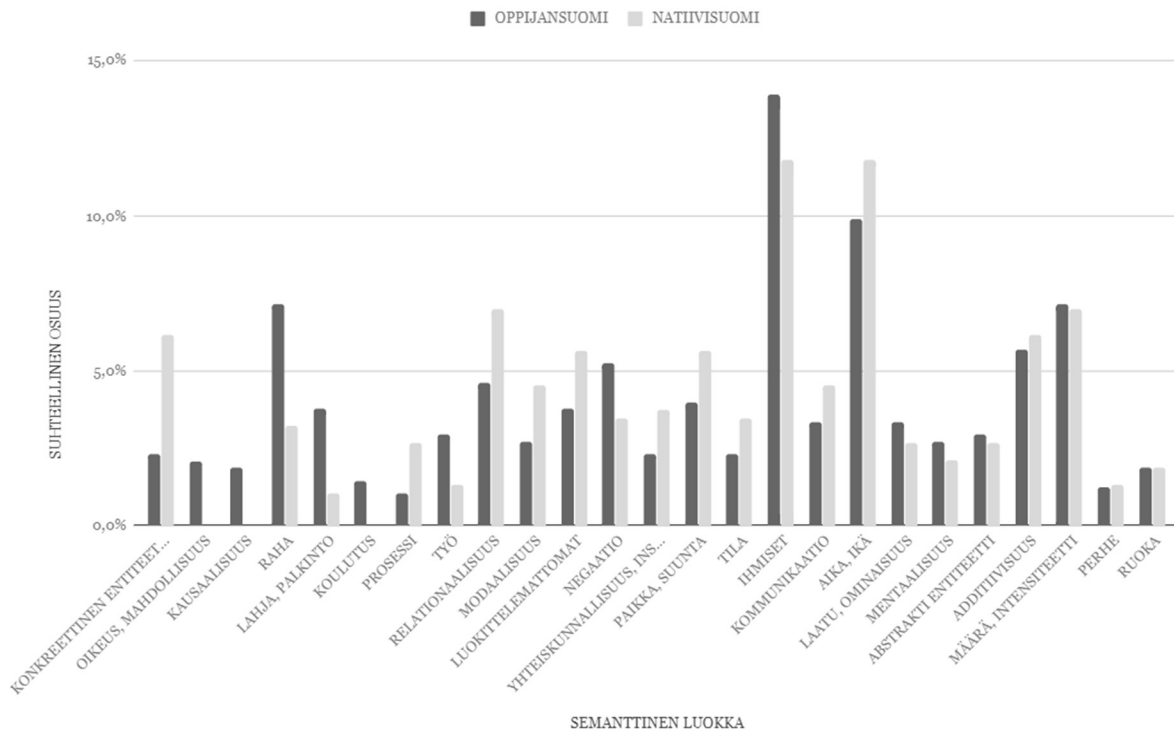
[205] Pienempi sai huuliharpun, joka parkaisi ilkeästi. (N)

[206] Lomakka sai siihenkin vastaukseksi vain lyhyen nyökkäyksen [-]. (N)

Oppijansuomessa ja natiivisuomessa SAADA-verbin merkitysluokassa 1 erot eivät ole kovin suuria: molemmissa taajimpien luokkien joukkoon kuuluvat 'ihmiset', 'aika, ikä' ja 'määrä, intensiteetti'. Molemmissa aineistoissa luokan 'ihmiset' primingia tuntuu selittävän SAADA-verbin argumenttien henkilötarkoitteinen luonne sekä merkitysluokka 1, jossa 'saaja' on osa merkitystä. Kun luokkia tarkastellaan tarkemmin laadullisesti, on kuitenkin havaittavissa sisällöllisiä eroja.

'Ihmiset'-luokka on hieman taajempi oppijansuomessa (13,9 %, n = 66) kuin natiivisuomessa (11,8 %, n = 44). Oppijansuomessa 'ihmiset'-luokan sanoista yli kolmannes on kolmannen persoonan pronomineja, kun taas natiivisuomessa lähes kolmannes koostuu nimistä (jotka toki myös ovat usein kolmannen persoonan viittauksia). Voisi spekuloida, johtuuko ero siitä, että oppijansuomessa henkilöviittauksia pyrittäisiin yksinkertaistamaan pronomini- viittauksilla. Luokassa 'aika, ikä' oppijansuomessa (9,9 %, n = 47) on suhteellisesti vähemmän sanoja kuin natiivisuomessa (11,8 %, n = 44). Oppijansuomessa korostuvat ilmaukset, jotka viittaavat menneisyyteen, tulevaisuuteen tai nykyisyyteen tai toistuvuuteen. Natiivisuomessa sen sijaan ei ole selkeitä alaluokkia, vaan ajalliset ilmaukset varioivat hyvin paljon. 'Määrä, intensiteetti' on oppijansuomessa (7,2 %, n = 34) vain aavistuksen taajempi kuin natiivisuomessa (7,0 %, n = 26). Oppijansuomessa 'määrä, intensiteetti' pääosin tarkkoina ja epätarkkoina lukumäärän ilmauksina, kun taas natiivisuomessa korostuvat kokoa ilmaisevat sanat.

Semanttisten luokkien väliset erot oppijan- ja natiivisuomen aineistojen välillä on kuvattu kuviossa 12 tilastollisen merkitsevyyden mukaisessa järjestyksessä ja tarkemmin z-arvoineen ja keskivirheineen liitteessä 11:



Kuvio 12. Oppijansuomen (475 sanetta) ja natiivisuomen (373 sanetta) semanttisten luokkien vertailu järjestettynä tilastollisen merkitsevyyden⁴⁷ mukaan.

Aineistojen välillä ei ole yhtään erittäin merkitsevää eroa, mutta luokat 'konkreettinen entiteetti'⁴⁸, 'oikeus, mahdollisuus'⁴⁹ ja 'kausaalisuus'⁵⁰ eroavat merkitsevästi. Melkein merkitsevästi eroavat luokat 'raha'⁵¹, 'lahja, palkinto'⁵² ja 'koulutus'⁵³. Lopuilla luokilla ero ei ole merkitsevä tai melkein merkitsevä⁵⁴: 'prosessi', 'työ', 'relaationaalisuus', 'modaalisuus', 'luokittelemattomat', 'negaatio', 'yhteiskunnallisuus, institutionaalisuus', 'paikka, suunta', 'tila', 'ihmiset', 'kommunikaatio', 'aika, ikä', 'laatu, ominaisuus', 'mentaalisuus', 'abstrakti entiteetti', 'additiivisuus', 'määrä, intensiteetti', 'perhe' ja 'ruoka'. Kaikki luokat ('oikeus, mahdollisuus', 'kausaalisuus' ja 'koulutus'), jotka esiintyvät oppijansuomessa, mutta eivät natiivisuomessa, ovat luonnollisesti korkeilla sijoilla erojen tilastollisessa merkitsevyydessä. Seuraavaksi tarkastellaan eniten eroavien semanttisten luokkien (sellaisten, jotka esiintyvät molemmissa aineistoissa) sisällöllisiä eroja aineistoittain: eroako muukin kuin taajuus?

⁴⁷ Raja-arvot: 1,96 melkein merkitsevä ($p < 0,05$); 2,57 merkitsevä ($p < 0,01$); 3,29 erittäin merkitsevä ($p < 0,001$).

⁴⁸ $z = -2,84$, $p < 0,01$.

⁴⁹ $z = 2,82$, $p < 0,01$.

⁵⁰ $z = 2,67$, $p < 0,01$.

⁵¹ $z = 2,51$, $p < 0,05$.

⁵² $z = 2,47$, $p < 0,05$.

⁵³ $z = 2,35$, $p < 0,05$.

⁵⁴ $z = -1,96 - 1,96$, $p > 0,05$.

Mielenkiintoista on, että eniten eroa on useammin natiivisuomessa kuin oppijansuomessa esiintyvä eli oppijansuomessa aliedustuva luokka 'konkreettinen entiteetti', joka ei ollut kummassakaan aineistossa taajimmasta päästä (oppijansuomessa sija 16/24 ja natiivisuomessa sija 5/21). Molemmissa aineistoissa 'konkreettinen entiteetti' pitää sisällään sekä elollisia (oppijansuomen aineistossa *kalan, koiran, koiranpennun ja kukkia*, natiivisuomen aineistossa *ajokoira-kin, hiireltä, petoa ja poroja*) että elottomia asioita (oppijansuomen aineistossa *arpalippua, avaimet, elokuva, kolmion, meripihkaa, sähköä ja tavaroita*, natiivisuomen aineistossa *fluoria, sähköä, ATK-järjestelmänsä, varastojen, avainta, paketin, narunpätkä, öljyä, saippuaa, lasivirtiriinikaappi, tuolinsa, maata, rasian ja välineistäni*). Vaikka natiivisuomen aineiston luokassa 'konkreettinen entiteetti' (n = 23) elolliset asiat esiintyvät suhteellisesti noin puolet harvemmin (17,4 %, n = 4) kuin oppijansuomen aineiston vastaavassa luokassa ('konkreettinen entiteetti' n = 11, josta 'elolliset asiat' 36,4 %, n = 4), ei ero ole tilastollisesti merkitsevä⁵⁵. Mielenkiintoista on, että AVAIN ja SÄHKÖ mainitaan molemmissa aineistoissa.

'Oikeus, mahdollisuus' ja 'kausaalisuus' -luokissa ei ole vertailtavaa, joten vertailussa siirrytään seuraavaksi merkitsevimpään luokkaan 'raha'. Kuten aiemmin oppijansuomen luokkien analysoinnissa kävi ilmi, oppijansuomessa 'raha' ilmenee suorana mainintana rahasta tai viittauksena yhteiskunnalta saataviin tukiin tai lainaan, maksettaviin veroihin, rahan ansaitsemiseen ja rahoitukseen. Natiivisuomessa eniten viittauksia rahan ansaitsemiseen (*lisätuloina, palkkaa, palkan ja rahapalkkaa*), rahaan itseensä (*rahaa x2 ja remonttirahat*), yhteiskunnallisiin raha-asioihin (*tullitaksan, veroista ja tuontitullitaksan*) ja yksittäistapauksina rahoitukseen (*apurahoja*) ja valuuttaan (*taalarin*). Natiivisuomesta siis puuttuu kokonaan yhteiskunnalta saatavat tuet, joista koostuu viidesosa (20,6 %, n = 7) oppijansuomen aineiston luokasta 'raha' (n = 34). Kun oppijansuomen aineistossa palkasta ja palkkiosta muodostuu 14,7 % (n = 5) 'raha'-luokasta (n = 34), natiivisuomen aineistossa osuus on suhteellisesti yli puolet enemmän, 33,3 % (n = 4), muttei kuitenkaan tilastollisesti merkitsevä⁵⁶. 'Raha'-luokassa on siis aineistoissa yhtäläisyyksiä ja eroa, mutta painopisteet ovat eri.

'Lahja, palkinto' -luokka on molemmissa aineistoissa pieni (oppijansuomessa 3,8 %, n = 18 ja natiivisuomessa 1,1 %, n = 4). Oppijansuomessa kaksi kolmasosaa, 66,7 % (n = 12) luokan sanoista on lahjoja (*joululahja, joululahjoja, lahjaksi x4, lahjan, lahjat ja lahjoja x4*). Loput ovat palkintoja (*Jussi-palkinnot ja Nobelin kirjallisuuspalkinnon*), stipendejä (*kansallis-stipendin x2*), perintöä (*perintönä*) ja bonuspisteitä (*bonuspisteitä*). Natiivisuomessa sen sijaan kolme neljäsosaa (75,0 %) luokan sanoista (n = 4) on palkintoja (n = 3: *palkinto, palkintojakin ja*

⁵⁵ z=1,29, p>0,05.

⁵⁶ z=-1,40, p>0,05.

kruunu), mutta sekaan mahtuu myös yksi perintö (*perintönä*). Ero palkintojen esiintyvyyden suhteen on tilastollisesti melkein merkitsevä⁵⁷, vaikka aineisto onkin hyvin minimaalinen. 'Lahja, palkinto' on viimeinen tilastollisten erojen perusteella käsiteltävä luokka, koska se on viimeinen tilastollista (melkein) merkitsevyyttä osoittava luokka, joka esiintyy molemmissa aineistoissa (seuraavaksi oleva 'koulutus' esiintyy vain oppijansuomen aineistossa).

6.4 Morfologisen ja semanttisen primingin korrelaatio oppijan- ja natiivisuomessa

Seuraavaksi neljänteen tutkimuskysymykseen vastataan siten, että aiemmin analysoitujen sananmuotojen ja semanttisten luokkien välistä korrelaatiota testataan tilastollisesti X^2 -riippuvuustestin avulla. Jotta X^2 -riippuvuustesti toimisi, ristiintaulukoinnissa 20,0 % arvoista saa olla pienempiä kuin viisi eikä yhtään alle 1:n arvoa saa olla (Heikkilä 2014: 13). Tämä aiheuttaa haasteita taulukolle, jossa on useita rivejä ja sarakkeita. Tämän vuoksi ristiintaulukoinnissa luokkia on jouduttu yhdistämään. Seuraavaksi käydään läpi luokkien yhdistämisen periaatteet, jonka jälkeen esitellään X^2 -riippuvuustestin tulokset. Luokat eivät yhdistyneet aivan saumattomasti, minkä vuoksi korrelaatiotuloksia tulee pitää suuntaa antavina. Tästä syystä X^2 -riippuvuustestin tulokset avataan myös laadullisesti.

6.4.1 Luokkien yhdistämisen periaatteet ristiintaulukoinnissa

Sananmuotoja on pyritty yhdistämään niin, että niillä olisi mahdollisimman laajasti yhdistävä tekijä. Molemmissa aineistoissa oliärkevintä yhdistää sananmuodot persoonan mukaisesti niin, että ensimmäisen, toisen, kolmannen ja passiivin muodot ovat yhdessä, mikäli mahdollista. Toinen vaihtoehto olisi ollut hyödyntää aikamuotoa, mutta ilman OLLA-verbiä tämä olisi ollut hankalaa. Näin ollen jako on tehty persoonittain. Molemmissa aineistossa on lisäksi luokka, johon on yhdistetty vähiten esiintyvät sananmuodot. Semanttisia luokkia on yhdistetty niin, että kapea-alaisia luokkia on yhdistetty sopiviin laajemman merkityksen luokkiin. Aina on siis pyritty löytämään yhdistettävien luokkien välille jokin yhteinen tekijä. Koska eri aineistossa taajuudet eroavat, eivät luokkayhdistelmät ole aivan identtiset, vaikka ne on tehty samoja periaatteita noudattamalla. Huomioitavaa on myös, että mikäli luokka on yksistään ollut tarpeeksi suuri, sitä ei ole yhdistetty muihin luokkiin eikä muita pienempiä luokkia ole yhdistetty

⁵⁷ $z=-2,18$, $p<0,05$.

siihen. Tämän vuoksi molemmissa aineistoissa on pienen frekvenssin luokat, vaikka jokin näistä luokista sopisikin muualle (esim. *saisi* ja *saadakseen* voisivat käydä luokkaan *saa, sai*, mutta tällöin *saanut* ja *saaneet* -luokka ei olisi tarpeeksi suuri). Luokat on siis yhdistämisestä huolimatta pyritty pitämään mahdollisimman tarkkoina, ja yhdistäminen on tehty vain pakosta.

Oppijansuomen aineistossa luokat on yhdistetty seuraavasti (yhdistetyn taajuusjärjestyksen mukaisesti):

sananmuodot

1. saa + sai
2. saada
3. saavat
4. sain + saamme + saan
5. saanut + saaneet + saisi + saadakseen

semanttiset luokat

1. abstrakti entiteetti + yhteiskunnallisuus, institutionalisuus + työ + koulutus + laatu, ominaisuus + oikeus, mahdollisuus
2. konkreettinen entiteetti + lahja, palkinto + raha + ruoka
3. ihmiset + perhe
4. negaatio + relationaalisuus + modaalisuus
5. määrä, intensiteetti + additiivisuus
6. aika, ikä
7. mentaalisuus + kommunikaatio
8. paikka, suunta + tila
9. luokittelemattomat + prosessi + kausaalisuus

Sananmuodoista luokkaa 1 yhdistää yksikön kolmannen persoonan indikatiivimuodot *saa* ja *sai*. Toisen luokan muodostaa yksin perusmuoto *saada* ja kolmannen *saavat*. Neljänteen luokkaan on yhdistetty ensimmäisen persoonan muodot yksiköstä ja monikosta: *saamme, saan* ja *sain*. Viimeinen luokka pitää sisällään loput sananmuodot *saanut, saaneet, saisi* ja *saadakseen*, jotka voidaan jakaa edelleen infiniittisiin muotoihin ja finiittiseen yksikön kolmatta persoonaa ilmaisevaan muotoon. Mikäli kolmannen persoonan muoto olisi yhdistetty luokkaan 1, olisi viimeinen luokka kuitenkin jäänyt liian pieneksi. Tämän vuoksi näistä pienistä luokista on koottu yksi suurempi luokka.

Semanttisten luokkien osalta ensimmäinen luokka koostuu 'abstraktista entiteetistä' sekä pienemmistä luokista, jotka myös voidaan mieltää kuuluvaksi 'abstraktiin entiteettiin' ('yhteiskunnallisuus, institutionaalisuus', 'työ', 'koulutus', 'laatu, ominaisuus' ja 'oikeus, mahdollisuus'). Vastaavasti toinen luokka muodostuu 'konkreettisesta entiteetistä' ja sille alisteisista luokista ('lahja, palkinto', 'raha' ja 'ruoka'). Kolmas luokka koostuu ihmisistä, eli luokista 'ihmiset' ja 'perhe'. Neljäs luokka pitää sisällään kieliopillis-semanttisiksi mielletävät luokat 'negaatio', 'relationaalisuus' ja 'modaalisuus'. Edelliseen luokkaan sopiva 'additiivisuus' yhdistyy kuitenkin luontevammin luokkaan viisi, jossa se sopii yhteen luokan 'määrä, intensiteetti' kanssa. Luokka kuusi muodostuu vain luokasta 'aika, ikä'. Seitsemännessä

luokassa yhdistyvät mentaaliseksi tai kognitiiviseksi mielletävät prosessit 'mentaalisuus' sekä 'kommunikaatio'. Kahdeksannessa luokassa yhdistyy sekä konkreettinen että abstrakti spatiaalisuus, 'paikka, suunta' ja 'tila'. Viimeisessä luokassa on 'luokittelemattomien' lisäksi muhin luokkiin sopimattomat 'prosessi' ja 'kausaalisuus'. Vaikka 'kausaalisuus' olisi periaatteessa sopinut muiden kieliopillisten ilmiöiden kanssa luokkaan neljä, olisi viimeinen luokka jäänyt vajaaksi ilman sitä – ja koska 'kausaalisuus' olisi ollut kieliopillisen luokan pienin ja sen vuoksi vähiten merkityksellinen luokka, päättyi se viimeiseen luokkaan.

Natiivisuomen aineistossa luokkajako on seuraava:

sananmuodot

1. sai + saa
2. saanut
3. saat + saatiin + saisi + saatu
4. saada
5. saivat + saavat
6. sain, saan

semanttiset luokat

1. aika, ikä + paikka, suunta + tila
2. abstrakti entiteetti + kommunikaatio + mentaalisuus + työ + laatu, ominaisuus + yhteiskunnallisuus, institutionaalisuus
3. negaatio + relationaalisuus + modaalisuus
4. määrä, intensiteetti + additiivisuus
5. ihmiset + perhe
6. konkreettinen entiteetti + lahja, palkinto + raha + ruoka
7. luokittelemattomat + prosessi

Natiivisuomen luokkien yhdistäminen osoittautui hankalammaksi kuin oppijansuomen, erityisesti sananmuototyyppien osalta. Ensimmäinen (*sai, saa*) luokka on sama kuin oppijansuomessa, mutta sen jälkeen on eroja. Toinen luokka *saanut* on erilainen sillä perusteella, että kyseinen sananmuoto esiintyy natiivisuomessa useammin kuin oppijansuomessa. Kolmas luokka on natiivisuomen pienempien taajuuksien sananmuodoista koostettu luokka. Luokka on olemassa siis sillä perusteella, että luokkien täytyy olla tarpeeksi suuria ja jakautua tarpeeksi tasaisesti, jotta X^2 -riippuvuustesti toimisi, ja tämä on ohjannut luokkien yhdistämistä (ja tehnyt siitä myös hyvin hankalaa). Loput sananmuototyyppien jaot ovat natiivisuomessa kuitenkin loogisia: luokka neljä sisältää yksistään perusmuodon *saada*, luokka viisi kolmannen persoonan monikon indikatiivissa *saivat, saavat* ja viimeinen luokka ensimmäisen persoonan yksikön indikatiivissa *sain, saan*.

Semanttisia luokkia on myös joutunut yhdistämään hieman eri tavalla. Merkittävin näistä on kenties ensimmäinen luokka, jossa on yhdistetty 'aika, ikä', 'paikka, suunta' ja 'tila'. Tämä johtuu siitä, että 'aika, ikä' yksistään ei täytä X^2 -riippuvuustestin ehtoja (näin olisi jäänyt 0-arvo ristiintaulukoinnin kohtaan *saisi* + 'aika, ikä') ja ainoa samankaltainen luokkayhdistelmä on 'paikka, suunta' ja 'tila', koska 'ajan' voi mieltää myös ajallisena paikkana. Kyse on siis kaikissa näissä luokissa jonkinlaisesta ulottuvuudesta. Toisen luokan yhdistävä tekijä on

abstraktisuus: 'abstrakti entiteetti', 'kommunikaatio', 'mentaalisuus', 'työ', 'laatu, ominaisuus' ja 'yhteiskunnallisuus, institutionaalisuus'. Toisin kuin oppijansuomessa, 'kommunikaatio' ja 'mentaalisuus' eivät riittäneet muodostamaan omaa luokkaansa, joten ne on yhdistetty muiden abstraktien asioiden kanssa. 'Kieliopillissemanttisten' luokkien (luokka 4), 'ihmisten' ja 'perheen' luokkien (luokka 5) ja 'konkreettisten entiteettien' (luokka 6) luokkien yhdistelmät ovat samanlaisia kuin oppijansuomessa. Myös viimeinen luokka on samanlainen lukuun ottamatta luokan 'kausaalisuus' puuttumista.

Luokkien yhdistämiseen on siis sovellettu samoja periaatteita, mutta aineisto ei kuitenkaan aina ole taipunut näihin periaatteisiin samoin tavoin. Tämän vuoksia aineistojen luokkien yhdistelyssä on pieniä eroja. Korrelaatioon pureudutaan tarkemmin seuraavassa alaluvussa siten, että tilastollisia tuloksia tarkastellaan laadullisella otteella. Tämä johtuu siitä, että luokkien sisältö täytyy ottaa laadullisesti huomioon, ennen kuin numeerista tietoa voi tulkita luotettavasti.

6.4.2 Korrelaatio morfologisen ja semanttisten primingin välillä oppijansuomessa

Oppijansuomessa korrelaatio sananmuotojen ja semanttisten luokkien välillä näyttäisi pelkän tilastollisen tarkastelun valossa olevan erittäin merkitsevää⁵⁸. Liitteessä 12 on tarkemmin aineiston ristiintaulukointi, odotusarvot ja kontribuutioarvot. Kuten aiemmassa alaluvussa tuli ilmi, täytyy tätä kuitenkin tulkita laadullisesti ja korrelaation voimakkuuteen suhtautua kriittisesti. X^2 -riippuvuudesta osoittaa, että oppijansuomessa korrelaatio olisi suurinta⁵⁹ seuraavien luokkien välillä (voimakkaimmasta korrelaatiosta heikompaan):

1. *saanut, saaneet, saisi, saadakseen* + 'luokittelemattomat', 'prosessi', 'kausaalisuus'⁶⁰ (+)⁶¹
2. *sain, saamme, saan* + 'aika, ikä'⁶² (+)
3. *saavat* + 'ihmiset', 'perhe'⁶³ (+)
4. *saada* + 'paikka, suunta', 'tila'⁶⁴ (+)

Suurin yhteys on sekä sananmuotojen että semanttisten luokkien pienimmistä luokista tai jämäluokista koostettujen luokkien välillä, eli sananmuotojen *saanut, saaneet, saisi, saadakseen*

⁵⁸ $X^2=81,31$, $p=0,00000358667489556$, vapausaste=32,00.

⁵⁹ Käsitteeseen on otettu mukaan sellaiset suhteet, joiden kontribuutioarvo on suurempi kuin 4,07, jolloin osuus X^2 -arvosta on vähintään 5,0 %. Absoluuttiset arvot on pyöristetty kahden desimaalin tarkkuudelle, prosentiosuudet yhden desimaalin tarkkuudelle.

⁶⁰ contributions-arvo=14,55.

⁶¹ (+) tarkoittaa myötäesiintymistä ja (-) välttelysuhdetta.

⁶² contributions-arvo=11,42.

⁶³ contributions-arvo=9,94.

⁶⁴ contributions-arvo=6,68.

(yht. $n = 70$) ja semanttisten luokkien 'luokittelemattomat', 'prosessi' ja 'kausaalisuus' (yht. $n = 32$) välillä. Molemmissa luokissa on paljon yksittäisiä elementtejä, eikä kumpikaan luokka muodosta kovin selkeää kokonaisuutta, joten tätä tulosta joudutaan tarkastelemaan purkamalla luokat pienempiin osiin. NUT-partisiippien (*saanut, saaneet*, yht. $n = 40$) kontekstissa luokat 'luokittelemattomat' (20,0 %, $n = 8$) ja 'ihmiset' (20,0 %, $n = 8$) nousevat suurimmiksi. 'Luokittelemattomat' koostuu tässä tapauksessa yksinomaan OLLA-verbin sananmuodoista, mikä on *saaneet*- ja *saanut*-sananmuodoille odotuksenmukaista, sillä kyseessä on liittotempusten toinen osa (perfekti ja pluskvamperfekti). *Saadakseen*-sananmuodolla ($n = 15$) ei ole havaittavissa selkeää merkitysympäristöä: suurin merkitysluokka *saadakseen*-muodolle on 'aika, ikä' (20,0 %, $n = 3$), mutta pienten määrien takia tähän kannattanee suhtautua varauksella. Sananmuodon *saisi* ($n = 15$) kontekstin suurin merkitysluokka on 'raha' (26,7 %, $n = 4$). *Saadakseen*- ja *saisi*-sananmuotojen konteksteissa on yhteensä vain yksi sana luokassa 'luokittelemattomat' ja kaksi sanaa luokassa 'kausaalisuus', mistä voi päätellä, että korrelaatio koskeekin vain sananmuotoja *saanut* ja *saaneet*. Luokassa 'luokittelemattomat' ($n = 18$) esiintyy yhdeksän eri sananmuoto-tyyppiä, mikä kertoo jonkinasteisesta variaatiosta, mutta suurin osa sananmuodoista on NUT-partisiippeja (44,4 %, $n = 8$). Luokassa 'kausaalisuus' ($n = 6$) jokainen sananmuoto on eri, ja luokassa 'prosessi' ($n = 5$) taajin sananmuoto esiintyy kahdesti. Tämän korrelaatiotuloksen kohdalla kyse on lähinnä siitä, että metodi paljasti liittotempuksen (oppijansuomen aineistoesimerkit 207–208) – merkityksen ja muodon välisestä suhteesta ei siis ole kyse.

[207] Yhden vuoden jälkeen molemmat olivat saaneet (*saaneet) kokoon aika paljon rahaa. (O)

[208] Vihdoinkin, Erkki on ostanut uuden kaivinkoneen, tai paremmin sanonut, hän on saanut rahaa pankilta lainaksi kaivinkoneen ostamisesta. (O)

Toiseksi vahvimman poikkeaman X^2 -riippuvuustesti osoittaa ensimmäisen persoonan sananmuotojen *sain, saamme, saan* (yht. $n = 65$) ja semanttisen luokan 'aika, ikä' ($n = 47$) välille. Oppijansuomen semanttisesta luokasta 'aika, ikä' sananmuotoja *saan, saamme* ja *sain* (luokassa 'aika, ikä' $n = 15$) on noin kolmannes (31,9 %). Sananmuodoista *saan, saamme* ja *sain* taajimmin semanttisessa luokassa 'aika, ikä' esiintyy *sain* ($n = 9$; kaikista luokan sananmuodoista 19,1 %, yhdistetyn luokan *saan, saamme* ja *sain* esiintymistä 60,0 %). Koko semanttisen luokan osalta yhtä usein esiintyy sananmuoto *saa* (kaikista luokan 'aika, ikä' sananmuodoista 19,1 %, $n = 9$), mitä X^2 -riippuvuustesti ei näyttänyt. Itse asiassa, kun kaikki kolmannen persoonan muodot lasketaan yhteen (*saa, saavat, sai* ja *saisi*), yhteismäärä on enemmän kuin ensimmäisten persoonan muodoissa (kolmannen persoonan muodot kattavat 44,7 % [$n = 21$] kaikista semanttisen luokan 'aika, ikä' sananmuodoista). Kun asiaa tarkastellaan toisin päin, niin

'aika, ikä' (n = 17) on taajin (23,6 %) semanttinen luokka sananmuotojen *saan*, *saamme* ja *sain* kontekstissa, minkä vuoksi yhteys näiden luokkien välillä todennäköisesti näkyy X^2 -riippuvuustestissä. Voi siis sanoa, että 'aika, ikä' -luokka suosii kolmatta persoonaa (*saa*, *saavat*, *sai* ja *saisi*) (oppijansuomen aineistoesimerkit 209–210), mutta ensimmäinen persoona (*saan*, *saamme* ja *sain*) suosii semanttista luokkaa 'aika, ikä' (oppijansuomen aineistoesimerkit 211–212).

[209] Eräänä päivänä hän saa kirjeen Anssille. (O)

[210] [– –] kissa ja Mies saavat maitoa toisinaan. (O)

[211] [– –] kun täytin yksitoista vuotta, sain lopulta kaksi marssu (*marsua). (O)

[212] Siellä saamme aina tarpeeksi ruokaa. (O)

Kolmas testin osoittama yhteys on sananmuodon *saavat* (n = 75) ja semanttisten luokkien 'ihmiset' ja 'perhe' (molemmat luokat yht. n = 72) välillä. Kun nämä semanttiset luokat yhdistää, saadaan 15,2 % kaikista oppijansuomen semanttisen luokittelun aineistosta. Näistä lähes kolmasosa (30,6 %, n = 22) esiintyy *saavat*-sananmuodon ympärillä. Sanoista luonnollisesti suurin osa on ryhmässä 'ihmiset' luokkien kokoeron vuoksi, mutta 3 sanaa (13,6 %) on ryhmästä 'perhe'. Vastaavasti kaikista *saavat*-sananmuodon esiintymistä kontekstissa esiintyy 'ihmiset' tai 'perhe' 29,3 %:n taajuudella. Voi siis sanoa, että oppijansuomessa sananmuoto *saavat* myötäesiintyy semanttisen luokan 'ihmiset' (ja 'perhe') kanssa (oppijansuomen aineistoesimerkit 213–214).

[213] Annikin lapset saavat kukin myös noin kymmenentuhannen markan. (O)

[214] Sää Suomessa on kaunis vielä ensi viikolla ja toivottavasti suomalaiset saavat makkaransa pian. (O)

Neljänneksi X^2 -riippuvuustesti osoittaa, että sananmuodon *saada* (n = 102) ja semanttisten luokkien 'paikka, suunta' ja 'tila' (molemmat luokat yht. n = 30) välillä olisi korrelaatiota. Luokkien 'paikka, suunta' ja 'tila' sananmuodoista 43,3 % (n = 13) onkin *saada*. *Saada*-sanamuoto on kuitenkin niin taaja, että semanttiset luokat 'paikka, suunta' ja 'tila' muodostavat yhteensä vain 12,7 % (n = 12) sen semanttisesta kontekstista. *Saada*-sananmuodon kontekstissa esiintyy yhteensä 22 eri semanttista luokkaa, eli lähes kaikki. Tämä kertoo siitä, että *saada*-sanamuodolla ei oikeastaan ole (ainakaan vahvaa) myötäesiintymispreferenssiä, mutta sen sijaan kontekstissa 'paikka, suunta' ja 'tila' *saada* on hyvin todennäköinen sananmuoto (oppijansuomen aineistoesimerkit 215–216).

[215] Miten sinne voidaan saada sähköä? (O)

[216] Hän oli kuullut että Norjassa voisi saada työn kalan perkaajana. (O)

6.4.3 Korrelaatio morfologisen ja semanttisten primingin välillä natiivisuomessa

Natiivisuomen aineistossa korrelaatio näyttää olevan tilastollisesti merkittävä⁶⁵, mutta kuten oppijansuomen aineistossa, myös natiivisuomen aineistossa numeeriseen tulokseen kannattaa suhtautua pienellä varauksella luokkien yhdistelyn vuoksi. Liitteessä 13 on kuvattu tarkemmin aineiston ristiintaulukointi, odotusarvot ja kontribuutioarvot. Natiivisuomessa X^2 -riippuvuustesti nostaa esille neljä suhdetta, jotka poikkeavat merkittävästi odotuksenmukaisista arvoista (järjestys merkittävyyden⁶⁶ mukaisesti):

1. *saanut* + 'luokittelemattomat', 'prosessi'⁶⁷ (+)
2. *saanut* + 'aika, ikä', 'paikka', 'tila'⁶⁸ (-)
3. *saivat, saavat* + 'määrä, intensiteetti', 'additiivisuus'⁶⁹ (+)
4. *saada* + 'negaatio', 'relaationaalisuus', 'modaalisuus'⁷⁰ (+)

Suurin poikkeama on sananmuodon *saanut* (n = 68) ja semanttisten luokkien 'luokittelemattomat' ja 'prosessi' (yht. n = 31) välillä. *Saanut*-sanamuodon kontekstissa 'luokittelemattomat' ja 'prosessi' kattavat 20,6 % (n = 14) rajatun kontekstin alueesta niin, että 'luokittelemattomat' kattaa 16,2 % (n = 11) ja 'prosessi' 4,4 % (n = 3). Kaikista 'luokittelemattomien' (n = 21) ja 'prosessien' (n = 10) yhteydessä esiintyvistä sanamuodoista *saanut* on 45,2 % (n = 14). 'Luokittelemattomista' osuus on 52,4 % (n = 11) ja 'prosesseista' 30,0 % (n = 3). 'Luokittelemattomien' osalta tuloksissa näkyy liittotempukset, sillä 'luokittelemattomat' koostuvat yksinomaan OLLA-verbin sanamuodoista apuverbin roolissa. Voinee siis sanoa, että poikkeama selittyy kieliopillisilla syillä luokkaan 'luokittelemattomat' (natiivisuomen aineistoesimerkit 217–219), mutta jonkinasteista taipumusta on semanttisella tasolla *saanut*-sanamuodon ja 'prosessi' -semanttisen luokan välillä (natiivisuomen aineistoesimerkit 218–219).

[217] Varjokuu **oli saanut** nimensä siitä, että sen pinta oli täynnä tummia varjomaisia läiskiä. (N)

[218] [– –] jos Hannibal **olisikin saanut** varastojen **täydennystä**... (N)

[219] Toisaalta on kohtuutonta ajatella, että kun työntekijä vuosia melulle altistavassa työssä **toimittuaan on saanut meluavamman, häneltä** sen vuoksi kielletään työnteko ja hänet pakotetaan ammatinvaihtoon.

⁶⁵ $X^2=46,02$, $p=0,031$, vapausaste 30,0.

⁶⁶ Käsitteelyyn on otettu mukaan sellaiset suhteet, joiden kontribuutioarvo on suurempi kuin 2,30, jolloin osuus X^2 -arvosta on vähintään 5,0 %. Absoluuttiset arvot on pyöristetty kahden desimaalin tarkkuudelle, prosenttiosuudet yhden desimaalin tarkkuudelle.

⁶⁷ contributions-arvo=12,33.

⁶⁸ contributions-arvo=3,80.

⁶⁹ contributions-arvo=3,61.

⁷⁰ contributions-arvo=3,03.

Sananmuoto *saanut* (n = 68) esiintyy myös seuraavassa poikkeamassa, jolloin tarkastelussa on sen suhde semanttisiin luokkiin 'aika, ikä', 'paikka, suunta' ja 'tila' (yht. n = 79). Edellinen poikkeama viittasi myötäesiintymään, tämä sen sijaan välttelysuhteeseen. Sananmuodon *saanut* kontekstissa 'aika, ikä' ja 'tila' ('paikka, suunta' ei lainkaan) kattavat yhdessä 10,3 % (n = 7), erikseen 'aika, ikä' 7,4 % (n = 5) ja 'tila' 2,9 % (n = 2). Yhteensä luokkien 'aika, ikä', 'paikka, suunta' ja 'tila' yhteydessä esiintyvistä sananmuodoista *saanut* kattaa 8,9 % (n = 7), erikseen luokasta 'aika, ikä' (n = 44) 11,4 % (n = 5), luokasta 'paikka, suunta' (n = 21) luonnollisesti 0,0 % ja luokasta 'tila' (n = 13) 15,4 %. Tämän perusteella voisi sanoa, että välttelysuhde on selkeä sananmuodon *saanut* ja semanttisen luokan 'paikka, suunta' välillä. Sen lisäksi *saanut*-sananmuodolla on useita semanttisia luokkia kontekstissaan, joten ei ole kovin todennäköistä, että se esiintyisi luokkien 'aika, ikä' ja 'tila' konteksteissa – joiden kontekstit sen sijaan ovat pienemmät, minkä vuoksi *saanut*-sananmuoto ei näytä niin vahvaa välttelypreferenssiä näissä konteksteissa.

Kolmas odotusten vastainen poikkeama on monikon kolmannen persoonan indikatiivin sananmuotojen *saivat*, *saavat* (yht. n = 43) ja semanttisten luokkien 'määrä, intensiteetti' ja 'additiivisuus' (yht. n = 48) välillä. Todellinen yhteisesiintyvyys on odotettua suurempi, eli kyse on myötäesiintymissuhteesta. Sananmuotojen *saivat* ja *saavat* konteksteissa 23,3 % (n = 10) on joko luokkaa 'määrä, intensiteetti' tai 'additiivisuus'. Vastaavasti näiden semanttisten luokkien yhteydessä esiintyvistä sananmuodoista 20,8 % (n = 10) on *saivat* tai *saavat*. Tarkemmin yksittäisiä sananmuotoja ja luokkia tarkastelemalla paljastuu, että suurin yhteys löytyy sananmuodon *saivat* (n = 28) ja sen ympäristössä esiintyvien luokkien 'määrä, intensiteetti' (n = 4) ja 'additiivisuus' (n = 5) välillä: 32,1 % (n = 9) *saivat*-sananmuodon kontekstista on joko 'määrä, intensiteetti' tai 'additiivisuus'. Vastaavasti yhteys näkyy myös toisin päin: *saivat* kattaa 18,8 % luokkien 'määrä, intensiteetti' ja 'additiivisuus' yhteydessä esiintyvistä sananmuodoista. Sen sijaan *saavat* (n = 15) näyttäisi välttelevän 'määrää' ja 'lisää' ilmaisevia luokkia, sillä *saavat*-sananmuodon kontekstissa esiintyy vain yksi 'additiivisuus'-luokan sana (6,7 %) eikä lainkaan 'määrä, intensiteetti' -luokan sanoja. Tästä voi päätellä, että *saivat*-sananmuodolla on yhteys määrälliseen merkitykseen (natiivisuomen aineistoesimerkit 220–221), kun taas *saavat*-muodolla ei (vaan suhde on jopa välttelevä). Onkin mielenkiintoista, että sananmuodot eroavat vain aikamuotomerkitykseltään, ja silti semanttisesta ympäristöstä nousee tämänkaltaisia eroja.

[220] Vuonna 1843 Uudenkaupungin tullikamari sai veneen ja Helsingin ja Turun tullikamarit saivat toiset partioveneensä. (N)

[221] Tavarat myytiin tullihuutokaupassa ja takavarikoijat saivat osuutensa [–]. (N)

Saada-sanamuodon (n = 64) ja kieliopillisten semanttisen ryhmien 'negaatio', 'relatiivisuus' ja 'modaalisuus' (yht. n = 56) välillä on myötäesiintymiseen viittaavaa poikkeama. Sanamuodon *saada* kontekstissa esiintyvistä merkitysluokista 23,4 % (n = 15) on joko 'negaatio', 'relatiivisuus' tai 'modaalisuus'. Tarkemman tarkastelun tuloksena paljastuu, että suurin osa näistä yhteisesiintymistä on luokan 'modaalisuus' kanssa, joka kattaa siis yksinään 20,3 % (n = 12) *saada*-sanamuodon kontekstista. Tämä ei oikeastaan ole ihme, sillä *saada* esiintyy usein modaalisten apuverbien kanssa (kuten tässä tapauksessa muun muassa sanojen *tulisi*, *voitu*, *voisin*, *pitäisi*, *täytyisi*, *pitää*, *saattoi* ja *taisi*). Yhteys luokkaan 'modaalisuus' on siis lähinnä kieliopillinen (natiivisuomen aineistoesimerkit 222–223). *Saada*-sanamuodon kontekstissa 'negaatio' esiintyy vain kerran (1,2 %) ja 'relatiivisuus' kahdesti (3,1 %), joten näiden luokkien osalta voidaan puhua jopa välttelysuhteesta.

[222] Fluoria tulisi saada usein ja pieninä annoksina [– –]. (N)

[223] Pitäisi kuulemma saada rahaa ratsastusleirille, kun muutkin pääsevät. (N)

6.4.4 Korrelaatioiden laadullinen vertailu oppijan- ja natiivisuomessa

Oppijansuomessa korrelaatio sananmuotojen ja semanttisten luokkien välillä näyttäisi olevan voimakkaampaa, mutta koska yhdistetyt luokat eivät olleet kauttaaltaan samoja kummassakin aineistossa, ei tätä voi sanoa aivan varmaksi. Tässä kohti on myös oleellista huomauttaa, että koska oppijansuomen ja natiivisuomen aineistoista muodostetut luokat ovat riippumattomuustestissä hieman erilaiset, tässä ei vertailla luokkia tai niiden numeerista dataa suoraan keskenään, vaan vertailu tapahtuu sen suhteen, mitä aineistojen sisäisesti havaitaan. Kyse on siis laadullisesta vertailusta.

Oppijansuomessa ensimmäisen persoonan muodoilla (*sain*, *saamme*, *saan*) on suurinta poikkeamista oletusarvoista (sananmuotojen *sain*, *saamme* ja *saan* yhteiskontribuutio $21,44/81,31 = 26,4\%$), kun taas natiivisuomessa ensimmäisen persoonan muodot (*sain*, *saan*) poikkeavat oletusarvoista vähiten (sananmuotojen *sain*, *saan* yhteiskontribuutio $1,91/46,02 = 4,2\%$). Kun tuloksia tarkastellaan semanttisten luokkien näkökulmasta huomataan, että oppijansuomessa 'luokittelemattomia' lukuun ottamatta 'ihmiset' ja 'perhe' aiheuttavat suurimman poikkeaman, kun taas natiivisuomessa 'luokittelemattomien' ja kieliopillisten luokkien jälkeen eniten poikkeamia oletusarvoihin aiheuttavat luokat 'aika, ikä', 'paikka' ja 'tila'. Natiivisuomessa 'ihmiset' ja 'perhe' aiheuttavat toisiksi vähiten poikkeamia 'konkreettisten entiteettien' yhdistelmäluokan jälkeen, mutta oppijansuomessa 'aika, ikä' on heti taajimman puhtaasti

semanttisten luokkien 'ihmiset' ja 'perhe' perässä. Seuraavaksi tarkastellaan riippumattomuustestin yksittäisiä merkittäviä suhteita.

Molemmissa aineistoissa esiin nousee vahva *saanut*-sanamuodon yhteys luokkaan 'luokittelemattomat' (oppijansuomessa 'luokittelemattomat', 'prosessi' [kontribuutio 14,55/81,31 = 17,9 %] ja 'kausalisuus', natiivisuomessa 'luokittelemattomat', 'prosessi' [kontribuutio 12,33/46,02 = 26,8 %]), jossa on OLLA-verbin kieliopilliset esiintymät. Tämä ei kerro kielivarianttien ominaispiirteistä, mutta osoittaa menetelmän nostavan esiin merkittäviä yhteyksiä. Toisaalta kummassakin aineistossa, mutta erityisesti natiivisuomen aineistosta, kyseinen suhde selittää suuren osan X^2 -arvosta, minkä vuoksi erityisesti natiivisuomen korrelaatiotulosta on syytä tarkastella kriittisesti.

Oppijansuomessa ensimmäisen persoonan muotojen ja 'aika, ikä' -luokan välillä on myötäesiintymistä (kontribuutio 11,42/81,31 = 14,0 %), mutta vastaava poikkeama natiivisuomen aineistossa on varsin pieni (kontribuutio 0,90/46,02 = 2,0 %). Oppijansuomessa sanamuodolla *saada* on myötäesiintymisen semanttisten luokkien 'paikka, suunta' ja 'tila' kanssa (kontribuutio 6,68/81,31 = 8,2 %). Natiivisuomessa sanamuodon *saada* suhde luokkiin 'paikka, suunta', 'tila' ja 'aika, ikä' sen sijaan on hyvin lähellä oletusarvoa (kontribuutio 0,44/46,02 = 1,0 %). Oppijansuomessa sanamuodon *saavat* ja semanttisten luokkien 'ihmiset' ja 'perhe' välillä on myötäesiintymistä (kontribuutio 9,94/81,31 = 12,2 %), mutta natiivisuomessa *saavat* ja *saivat* -sanamuotojen ja luokkien 'ihmiset' ja 'perhe' yhteisesiintymisen ei juuri poikkea oletusarvoista (kontribuutio 0,98/46,02 = 2,1 %).

Natiivisuomen aineistossa sanamuoto *saanut* näyttäisi välttelevän luokkia 'aika, ikä', 'paikka, suunta' ja 'tila' (kontribuutio 3,80/46,02 = 8,3 %). Tarkemman analyysin tuloksena paljastuu, että välttelysuhde koskee erityisesti luokkaa 'paikka, suunta'. Oppijansuomessa sanamuotojen *saanut*, *saaneet*, *saisi* ja *saadakseen* sekä semanttisten luokkien 'paikka, suunta' ja 'tila' välillä ei ole poikkeamaa oletusarvoista (kontribuutio 0,08/81,31 = 0,0 %).

7 PÄÄTÄNTÖ

7.1 Tutkimustulosten yhteenveto

Ensimmäinen tutkimuskysymys koskee SAADA-verbin taajimpia merkitystä oppijan- ja natiivikielessä. ICLFI:stä kerätyssä oppijankielen aineistossa (n = 513) taajin merkitysluokka on Kielitoimiston sanakirjan merkitys 1 ‘tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkun tai jonkin omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jokin tulee’. Tämä merkitysluokka kattaa koko oppijansuomen aineistosta 28,1 % (n = 144). Käännössuomen korpuksen alkusuomen aineistosta kerätyssä natiivisuomen aineistossa (n = 813) taajin merkitysluokka on Kielitoimiston sanakirjan merkitys 8 ‘onnistua, kyetä tekemään t. aiheuttamaan jtk t. saattamaan tekeminen päätökseen”, joka kattaa 24,5 % (n = 199) natiivisuomen aineistosta. Oppijansuomen aineistossa natiivisuomen taajin luokka 8 kattaa vain 5,3 % (n = 27) ja natiivisuomessa oppijansuomen taajin luokka 1 kattaa 15,1 % (n = 123). Kun aineistojen eri merkityksien osuuksia tarkastellaan tilastollisesti, sekä luokkien 1 ja 8 kohdalla että luokkien 5 ‘jnk tulemista jkn aistimusten, tiedon t. ymmärryksen piiriin’ ja 3 ‘ilmaantua tai tulla jonkin osaksi tai hyväksi’ kyse on tilastollisesti erittäin merkitsevästä eroista. Voidaan siis sanoa, että SAADA-verbin polyseeminen käyttö eroaa oppijan- ja natiivisuomessa. Oppijansuomessa aliedustuvat SAADA-verbin merkitykset 8 ja 3 ja yliedustuvat merkitykset 1 ja 5.

Toisella tutkimuskysymyksellä selvitetään, millaista morfologista primingia oppijankielen SAADA-verbin taajimmalla ja natiivikielen vastaavalla merkityksellä (eli merkityksellä 1 ‘tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkun tai jonkin omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jokin tulee’) esiintyy. Oppijansuomen aineiston taajimman merkitysluokan (n = 144) taajimmat SAADA-verbin sananmuodot ovat *saa* (19,4 %, n = 28) ja *saada* (19,4 %, n = 28). Yhteensä oppijansuomen SAADA-verbin taajimmassa merkitysluokassa eri sananmuototyyppinä on 21. Natiivisuomen vastaavan merkitysluokan (n = 123) taajimmat SAADA-verbin sananmuodot ovat *saanut* (14,6 %, n = 18) ja *sai* (14,6 %, n = 18). Yhteensä natiivisuomen merkitysluokassa 1 eri sananmuototyyppinä on 29. Tilastollisesti erittäin merkitsevästi eroaa sananmuoto *saat*, joka kattaa natiivisuomen sananmuodoista 7,3 % (n = 9). Laadullinen sananmuototyyppien tarkastelu paljastaa, että oppijansuomessa finiittisiä sananmuototyyppinä on enemmän kuin infiniittisiä ja natiivisuomessa infiniittisiä tyyppinä on enemmän kuin finiittisiä. Lisäksi natiivisuomen

finiittisistä muodoista suhteellisesti suurempi osa on imperfektissä kuin oppijansuomessa. Myös infiniittisten muotojen runsaus kieli osaltaan menneiden aikamuotojen runsaammasta käytöstä natiivisuomessa. Molemmissa kielivarianteissa kolmannen persoonan viittaukset ovat vahvoja, mikä voi tosin johtua myös kirjoitetun tekstin konventioista. Tulosten perusteella voi sanoa, että saman merkityksen sisällä oppijansuomessa ja natiivisuomessa morfologinen priming eroaa tilastollisesti, mutta myös laadullisesti. Oppijansuomessa aliedustuvat muodot *saat, saanut, saivat* ja *saatu* ja yliedustuvat muodot *saa, saavat* ja *saamme*. Lisäksi tuloksista voi odotustenmukaisesti vahvistaa vallalla ollutta käsitystä, että natiivisuomessa on suurempaa morfologista variaatiota.

Kolmas tutkimuskysymys hyödyntää myös merkitysluokan 1 aineistoa: tarkoitus on selvittää, millaista semanttista primingia oppijankielen SAADA-verbin taajimmalla ja natiivikielen vastaavalla merkityksellä esiintyy. Oppijansuomen merkitysluokassa 1 (n = 144) semanttisia luokkia esiintyy 24 luokittelemattomat pois luettuna. Taajimmat semanttiset luokat ovat 'ihmiset' (13,9 %, n = 66), 'aika, ikä' (9,9 %, n = 47), 'määrä, intensiteetti' (7,2 %, n = 34) ja 'raha' (7,2 %, n = 34). Natiivisuomen merkitysluokassa 1 (n = 123) semanttisia luokkia esiintyy 21 luokittelemat pois luettuna. Taajimmat semanttiset luokat ovat 'ihmiset' (11,8 %, n = 44), 'aika, ikä' (11,8 %, n = 44), 'määrä, intensiteetti' (7,0 %, n = 26) ja 'relaationaalisuus' (7,0 %, n = 26). Semanttisia luokkia on molemmissa aineistoissa siis melko paljon: etenkin pieniä, mutta merkitykseltään selkeitä luokkia. Tämä johtunee siitä, että SAADA on hyvin monikäyttöinen ja yleinen verbi, minkä vuoksi sille ei nouse yhtä tai muutamaa selkeää semanttista primingia. Eroja taajimpien semanttisten luokkien välillä on hyvin vähän.

Laadullisesti tarkasteltuna muutamia eroja on kuitenkin löydettävissä. Oppijansuomessa 'ihmiset'-luokan sanat ovat todennäköisesti kolmansien persoonien pronomineja, kun taas natiivisuomessa luokan 'ihmiset' sanat ovat todennäköisesti nimiä. 'Aika, ikä' -luokan sanat ovat suhteessa johonkin muuhun aikaan (menneisyys, nykyisyys tai tulevaisuus) tai ilmaisevat toistuvuutta. Natiivisuomessa selkeää sisäistä merkitystä luokassa 'aika, ikä' ei ollut. 'Määrä, intensiteetti' -luokka ilmenee oppijansuomessa täsmällisinä tai epätäsmällisinä lukumäärän ilmauksina, kun taas natiivisuomessa luokka keskittyy enemmän koon ilmaisemiseen.

Semanttisten luokkien suhteellisissa osuuksissa ei ole aineistojen välillä yhtään erittäin merkitsevää eroa, mutta luokat 'konkreettinen entiteetti', 'oikeus, mahdollisuus' ja 'kausaalisuus' eroavat merkitsevästi. Näistä luokista 'konkreettinen entiteetti' esiintyy merkittävästi useammin natiiviaineistossa, ja luokan sisällä elollisiin asioihin viittaavia sanoja on noin puolet harvemmin kuin oppijansuomen aineistossa. 'Oikeus, mahdollisuus' ja 'kausaalisuus' esiintyvät vain oppijansuomen aineistossa. Melkein tilastollisesti merkitsevästi käytöltään eroavan

luokan 'raha' suhteen natiivisuomesta puuttuu kokonaan yhteiskunnalta saataviin tukiin viittaavat sanat, kun taas oppijansuomen aineistossa nämä muodostavat viidesosan. Tulokset kertovat siitä, että SAADA-verbin merkitysluokan 1 kotekstissa esiintyvällä semanttisella primingilla ei ole juuri eroa. Kuitenkin, kun semanttisia luokkia tarkastelee laadullisesti, eroja on löydettävissä. Oppijansuomessa aliedustuu luokka 'konkreettinen entiteetti', kun taas luokat 'raha' ja 'lahja, palkinto' yliedustuvat. Tuloksista voi päätellä myös, että verbin merkitys on vahvasti yhteydessä sen semanttiseen ympäristöön: polysemiaa voisikin luonnehtia paradigmaattisena semanttisena primingina, joka on yhteneväinen myös syntagmaattiseen semanttiseen primingiin.

Neljännän tutkimuskysymyksen funktiona on selvittää morfologisen ja semanttisen primingin välistä korrelaatiota saman SAADA-verbin merkitysluokan sisältä. X^2 -riippumattomuustesti paljastaa, että oppijansuomessa näyttäisi olevan vahvempi sisäinen korrelaatio muodon ja merkityksen välillä kuin natiivisuomessa. Tarkempi tarkastelu paljastaa, että korrelaation vahvuuden lisäksi myös korreloivat tai toisiaan välttelevät muuttujat vaihtelevat eikä samankaltaisuuksia löydy lukuun ottamatta *saanut*-muodon yhteyttä 'luokittelemattomiin' (OLLA-apuverbiin), mikä johtunee kieliopillisesta rakenteesta eikä kielivarianttien ominaisuuksista.

Oppijansuomessa myötäesiintymistä löytyy siten, että 'aika, ikä' -luokka suosii kolmannen persoonan sananmuotoa, mutta ensimmäisen persoonan sananmuodot suosivat semanttista luokkaa 'aika, ikä'. Sananmuoto *saavat* ja semanttinen luokka 'ihmiset' korreloivat keskenään, ja lisäksi semanttisilla luokilla 'paikka, suunta' ja 'tila' on taipumusta esiintyä sananmuodon *saada* kanssa. Natiivisuomessa sananmuoto *saanut* on välttelysuhteessa ulottuvuutta ilmaisevien semanttisten luokkien, 'aika, ikä', 'paikka, suunta' ja 'tila' kanssa. *Saivat* myötäesiintyy luokkien 'määrä, intensiivisyys' ja 'additiivisuus' kanssa, kun taas *saavat* välttelee niitä. Lisäksi sananmuodolla *saada* on kieliopillinen yhteys luokkaan 'modaalisuus' ja välttelysuhde luokkiin 'negaatio' ja 'relationaalisuus'.

7.2 Tulokset osana traditiota

Tutkimus sijoittuu jatkumoon kontekstuaalisen semantiikan ja fraseologian tutkimusten kanssa vahvistaen käsityksiä kontekstin tärkeydestä ja vakiintuneista merkityksistä. Tutkimuksessa vahvistettiin myös leksikaalisen primingin teoriaa uudella tavalla, nimittäin kahden erilaisen fraseologisen yksikön väliltä. Tiedämme nyt, että erilaisten yksiköiden välillä voi vallita tilastollinen yhteys. Tämä tulos haastaa myös fraseologisen yksikön määritelmän: olisiko

mahdollista, että yksikön määritelmää vietäisiin pidemmälle siten, että lekseemi + ilmiö voisi olla [lekseemi + ilmiö] + [lekseemi + ilmiö]? Tämän ajatuksen vahvistaminen tarvitsee lisätutkimusta, mutta toivottavasti se herättelisi ajatuksia tutkimusyhteisössä.

Tutkimus myös täydensi kuvaa itsenäisten sananmuotojen luonteesta. Sinclair (1991) on esittänyt, että sananmuotoja pitäisi käsitellä itsenäisinä fraseologisina yksiköinä ja Stubbs (2001) on jo vahvistanut eri sananmuotojen kollokoivan eri tavoin, mutta tässä tutkimuksessa sananmuotojen ja semanttisten luokkien tilastollinen analyysi paljasti myös eri muotojen myötäesiintyvän tai olevan välttelysuhteessa keskenään. Erilaisilla sananmuodoilla on siis taipumuksia esiintyä erilaisissa semanttisissa ympäristöissä – jopa verbin saman merkityksen sisällä. Sananmuotoja on tutkittu melko vähän itsenäisinä yksikköinä, joten tämän tutkimuksen tuloksilla on painoarvoa – tutkimus osoittaa, että sananmuotojen syntagmaattisilla suhteilla on eroa, joten vahvistan väitteen, ettei fraseologiaa yksiköitä tulisi tarkastella lemmatasolla (ks. Sinclair 1991; Stubbs 2001). Tutkimus otti myös käytännössä askeleen polyseemisuuden huomioon ottamista fraseologisessa tutkimuksessa, vaikka eri merkitysten fraseologisten piirteiden vertailu ei mahtunutkaan tähän tutkimukseen mukaan. Tutkimus havainnollisti Jantusen ja Brunnin (2012: 87) toteamuksen, jonka mukaan verbin täydennykset ja sen erilaiset merkitykset ovat yhteydessä toisiinsa (VISK § 449), joten kun tutkitaan verbin kanssa myötäesiintyvien elementtien morfologisia piirteitä, tutkitaan välttämättä myös verbin eri merkityksiä. Eri sananmuotojen tutkimisen lisäksi olisi siis hyvä rajata leksikaalinen yksikkö tiettyyn merkitykseen, jotta sanoista ja niiden välisistä suhteista saatava kuva olisi mahdollisimman tarkka.

Oppijankielen fraseologisen tutkimuksen osalta tulokset asettuvat yhtenäiseen linjaan tradition kanssa: oppijansuomi eroaa natiivisuomesta fraseologisten piirteiden osalta. Vähemmän huomiota on saanut sanojen polyseeminen käyttö oppijansuomessa ja natiivisuomessa (ks. kuitenkin Jantunen & Brunni 2012), minkä vuoksi tulos polyseemisen käytön erosta on tärkeä. Tutkimus antaa uutta näkökulmaa oppijansuomen universaaleihin, erityisesti oppijankielen yksinkertaisuuteen ja kysymykseen siitä, *kuinka* oppijankieli on natiivisuomea yksinkertaisempaa. Vahvan korrelaation muodon ja merkityksen välillä voi nähdä variaation vähäisyytenä (vrt. heikompi korrelaatio natiivisuomen aineistossa). Voisikin spekuloida, johtuuko muodon ja merkityksen natiivisuomea vahvempi korrelaatio liiallisesta tukeutumisesta valmiiksi opeteltuihin fraaseihin puolivalmiiden konstruktioiden sijaan (ks. Sinclair 1991) tai onko sillä merkitystä, että kielenopetuksessa käsitellään usein tietty rakenne ja teema yhdessä (esim. ruoka ja partitiivi). Tämän vahvistaminen tarvitsee kuitenkin lisätutkimusta. Kuuluvaisen (2015: 57–58) mukaan fraseologiset virheet näyttävätkin liittyvän tavalla tai toisella oppijankielen

universaaleihin, joten kenties natiivikielestä poikkeava korrelaatio merkityksen ja muodon väliltä voisi myös linkittyä niihin.

Tilastollinen lähestymistapa muodon ja merkityksen väliseen tutkimukseen tuo tutkimuskentälle uudenlaisen näkökulman tarkastella merkityksen muodostumista ja sanaston käyttäytymistä. Menetelmä tarjoaa mahdollisuuksia laajentaa fraseologisen yksikön käsitettä. Tämän tutkimuksen oli tavoitteena kokeilla uudenlaista menetelmää fraseologiseen tutkimukseen – uskallankin sanoa, että pienellä menetelmän viilaamisella sillä on mahdollista tutkia mitä vain kielellistä ilmiötä.

Vaikka tutkimuksessa tutkittiinkin laajempia ilmiöitä SAADA-verbin avulla, saatiin SAADA-verbistäkin tarkempi kuvaus: sen polyseemisestä käyttäytymisestä, morfologisesta primingista, semanttisesta primingista sekä morfologisen ja semanttisen primingin välisestä yhteydestä. Semanttisen primingin osalta yhtäläisyyksiä löytyi ConLexiksen mainitsemiin SAADA-verbin kollokaatteihin, sillä AIKA ja RAHA esiintyvät molemmissa aineistoissa semanttisina luokkina. Muiden kollokaattien ja tämän tutkimuksen semanttisten luokkien eroja voi selittää se, että tässä tutkimuksessa semanttisen primingin analyysi tehtiin vain yhdelle verbin merkitykselle eikä koko lemmalle. Tämä huomio tukee teoriaa, jonka mukaan lemman eri merkityksillä on erilainen konteksti – ja minkä vuoksi polysemiaa ei tulisi sivuuttaa leksikon kuvauksessa.

7.3 Tutkimuksen onnistumisen arviointi

Tämän maisterintutkielman perimmäisinä tavoitteina oli selvittää muodon ja merkityksen välistä suhdetta oppijankielessä ja natiivikielessä ja luoda tutkielman laajuuden puitteissa mahdollisimman syvälinen kuva oppijankielen ja natiivikielen fraseologisista eroista. Tutkimusasetelma tuntui toimivan tavoitteisiin nähden hyvin. Tässä tutkimuksessa on pyritty mahdollisimman tarkasti ja eettisesti kuvamaan käytetyt aineistot, ohjelmistot versionumeroineen, aineistonkeruumetodit ja tutkimusmenetelmät, jotta tutkimusprosessi olisi mahdollisimman läpinäkyvä ja siten tarkasti toistettavissa.

Pienessä aineistossa korostuu kuitenkin tehtävänantojen ja jopa opiskelupaikkakunnasta kumpuavat kielenaineokset (Jantunen 2007: 79), minkä vuoksi tarvittaisiin laajempaa tutkimusta. Aineistojen pieni koko aiheutui ongelmaksi myös X^2 -riippumattomuustestin osalta, minkä vuoksi luokkia täytyi yhdistellä eri tavoin. Tämä vaikeutti korrelaatiotulosten vertailua aineistojen välillä. Menetelmän ongelmat olivat kuitenkin tiedossa koko prosessin ajan, joten

ongelmia pystyi paikkaamaan laadullisella tarkastelulla. Riippumattomuustesti pystyi kuitenkin kiistattomasti osoittamaan mielenkiintoiseen yhteyteen luokkien yhdistelystä huolimatta. Pie-nistä ongelmista huolimatta kokeellinen lähestymistapa muodon ja merkityksen välisen suhteen tarkasteluun tuotti kuitenkin rohkaisevia tuloksia, joiden perusteella on paikallaan kehittää me-netelmää tarkemmin laajemmassa tutkimuksessa.

Vaikka tutkimusasetelma osoitti, että verbin sisäisen merkityksen ja semanttisen ympä-ristön välillä vallitsee yhteys, voi asetelmaa pitää osin itseään ennustavana. SAADA-verbin merkitysluokassa 1 on kyse 'tapahtumasta jossa jokin esine, hyödyke, oikeus tms. siirtyy jonkin tai jonkun omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jokin tulee', joten molempien aineistojen kontekstissa taajimpana esiintynyt mer-kitysluokka 'ihmiset' ei liene yllätys, koska 'saaja' on jo merkitysluokan kriteereissä vahvasti läsnä. Myös sellaisten luokkien kuin 'raha', 'työ' ja 'lahja, palkinto' nouseminen selittynee verbin sisäisellä merkityksellä, sillä rahan tai työn saaminen kuuluvat 1. merkitysluokkaan. Oi-keastaan voisi varovaisesti sanoa, että SAADA-verbin merkitys itsessään määrääkin jossain määrin sen semanttisen primingin. Tämän todentamista varten tutkimusasetelma pitäisi toisin-taa saman verbin useilla eri merkityksillä.

Lisäksi voi pohtia, kuinka paljon aineistonkeruumenetelmä vaikutti lopulliseen aineis-toon. Systemaattinen satunnaisotanta ei ota huomioon eri genrejen suhteellisia osuuksia. Oppi-jansuomen korpuksessa tietyt tehtävät saattavat olla yliedustettuina ja natiivisuomen aineistossa tietty genre saattaa dominoida aineiston koostumusta. Koska jokaisella genrellä on omat kon-ventionsa, saattaa aineiston mahdollinen vinoutuminen vaikuttaa tuloksiin.

7.4 Tutkimustulosten soveltamisala ja jatkotutkimus

Nyt kun fraseologiasta ja leksikaalisesta yksiköstä on saatu uutta tietoa, on aika pohtia, mitä tiedolla voi tehdä. Pitkän tähtäimen tavoitteena on saada tutkimustieto osaksi opetusta siten, että suomenoppijat hyötyisivät tutkimuksesta. Tällä tutkimuksella on siis pedagoginen intressi, vaikka analyysin kohteena onkin ollut valmis tuotos oppimisprosessin sijaan. Sanaston frase-ologiselle opetukselle onkin hyvin paljon haasteita ja avoimia kysymyksiä, eikä korpusten ole-massaolo ei riitä siihen, että niiden sisältämä tieto siirtyisi opetukseen. (Jantunen 2009: 373.)

Pitäisikin pohtia, missä vaiheessa sanastoa aletaan opettaa yhtä sanaa laajempina koko-naisuuksia (ks. Jantunen 2009: 373). Tässä tutkimuksessa keskityttiin edistyneisiin kielenoppi-joihin, ja fraseologista eroavuutta natiiveihin oli edelleen paljon. Tästä herääkin kysymys,

pitäisikö oppijat totuttaa fraseologiaan jo kielenoppimisen alkeista asti. Ongelmana kuitenkin on, että taajaan käytetyillä sanoilla on useita merkityksiä ja käyttötapoja, mikä tekee niiden kuvaamisesta haasteellista (mp.). Tämä tutkimus on omalta osaltaan laajentanut tietämystämme taajaan käytetystä SAADA-verbistä usealla fraseologisella tasolla. S2-opetuksessa tutkimuksen antia voi hyödyntää SAADA-verbin kuvauksen osalta molemmissa kielivarianteissa, koska tutkimus osoitti kirjoitetun kielen osalta, kuinka oppijat käyttävät verbiä eri tavoin kuin natiivit. Näiden erojen tiedostaminen voi johtaa lopulta erojen kaventumiseen.

Opetuksen tueksi olisikin saatava enemmän todelliseen kielenkäyttöön perustuvaa materiaalia, sillä oppikirjojen kieli saattaa erota paljonkin todellisesta kielenkäytöstä (ks. Jantunen 2009b: 374). Tällä tutkimuksella on oppimateriaaleille kahdenlaista antia: a) suorat tutkimustulokset SAADA-verbin kuvaukseen ja tyypilliseen käyttöön tietyn merkityksen osalta ja b) menetelmällinen ja tutkimusasetelmallinen anti, jolla vastaavaa tietoa voi tuottaa lähes mistä vain tarvittavasta kielen ilmiöstä sopivan aineiston löytyessä. En näe estettä, etteikö opiskelijoillekin (ryhmästä riippuen toki) voisi antaa (valmiiksi tehtyjä) konkordanssirivejä tutkittavaksi ja saada lisää tietoa sitä kautta.

Leksikkotasolla jatkotutkimusta tarvittaisiin selvittämään, mistä tässä tutkimuksessa havaitut erot ja samankaltaisuudet SAADA-verbin fraseologisessa käytössä johtuvat. Tämän tutkimuksen valossa havaintoja voi spekuloida ja hahmotella hypoteeseja ja jatkotutkimusasetelmia, mutta mitään varmaa syistä ei voi sanoa. Lisäksi jatkotutkimusta tarvittaisiin laajentamaan tämän tutkimuksen antia muihinkin kielen tasoille. Tarvittaisiin vastaavaa fraseologista käyttöä koskevaa tietoa muista verbeistä, muista sanaluokista ja muista ilmauksista – tämän tutkimuksen tutkimusasetelmaa voi soveltaa mihin tahansa suomen kielen ilmaukseen. Lisäksi sanan useamman sisäisen merkitysluokan (ja sen sisällä tehtyjen muiden analyysien, kuten tässä tutkimuksessa tarkasteltujen morfologisen ja semanttisen primingin) vertailu olisi oleellista, jotta muodon ja merkityksen välisen suhteen tutkimus täydentyisi.

Edelleenkin ei oppijansuomesta (tai juuri muitakaan oppijankielistä) ole kovin montaa puhutun formaatin korpusta. Puhutun kielen fraseologinen tutkiminen olisi kuitenkin yhtä tärkeää kuin kirjoitetun, sillä vain kirjoitetun kielen korpuksia hyödyntämällä kokonaiskuva oppijankielestä vääristyy. Tutkimuksen laajentaminen puhuttuun kieleen olisi ensiarvoisen tärkeää myös siksi, että esimerkiksi maahanmuuttajille puhuttu kieli näyttäytyy arjessa aivan eri merkityksessä kuin suomea ulkomailla opiskeleville. Videomuotoisia korpuksia käyttämällä (annotaation avulla) tämän tutkimuksen menetelmin voisi tutkia jopa kielellisten elementtien sekä eleiden ja ilmeiden välistä yhteyttä ja näin laajentaa fraseologinen tutkimus pelkästä kielellisestä kotekstista todelliseen firthiläiseen tilannekontekstiin.

LÄHTEET

Lähteet

- Anthony, Laurence 2018: *AntConc Windows 64-bit 3.5.7*. – <http://www.laurenceanthony.net/software/antconcl/27.5.2018>
- Arppe, Antti 2002: The usage patterns and selectional preferences of synonyms in a morphologically rich language. – Anne Morin & Pascal Sélobiot (eds.), *Proceedings of the 6th international conference on textual data statistical analysis (JADT 2002) Vol. 1* s. 1–21. Rennes: INARIA.
- Barnbrook, Geoff 1996: *Language and computers. A practical introduction to the computer analysis of language*. Edinburgh Textbooks in Empirical Linguistics. Edinburgh: Edinburgh University Press.
- Bley-Vroman, Robert 1983: The comparative fallacy in interlanguage studies: the case of systematicity. – *Language Learning* 33 s. 1–17.
- Borin, Lars, Forsberg, Markus & Roxendal, Johan 2012: *Korp – the corpus infrastructure of Språkbanken*.
- Brazil, David 1995: *A grammar of speech*. Oxford: Oxford University Press.
- Bundesministerium für Bildung und Forschung 2018: *Bundesausbildungsförderungsgesetz – BAföG*. – <https://www.xn--bafg-7qa.de/de/bundesausbildungs-foerderungsgesetz---bafog-204.php> 10.6.2018.
- Cruse, Alan 1986: *Lexical semantics*. Cambridge: Cambridge University Press.
- 2000: *Meaning in language: an introduction to semantics and pragmatics*. Oxford: Oxford University Press.
- Corder, Pit 1976: *Miten kielitiedettä sovelletaan* (suom. Maria Vilkuna). Helsinki: Suomalaisen kirjallisuuden seura.
- Cowie, Anthony 1994: Phraseology. – R. E. Asher (ed.), *The encyclopedia of language and linguistics vol. 6* s. 3168–3171. Oxford: Pergamon Press.
- 1998: Introduction. – Anthony P. Cowie (ed.), *Phraseology: Theory, analysis and applications* s. 1–20. Oxford: Oxford University Press.
- Cutting, Joan 2000: Written errors of international students and English native speaker students. – George M. Blue, James Milton & Jane Saville (eds.), *Assessing English for academic purposes*. Frankfurt am Main: Peter Lang.
- Dagneaux, Estelle, Denness, Sharon & Granger, Sylviane 1998: Computer-aided Error Analysis. – *System* 26 (2) s. 471–490.
- Dewaele, Jean-Marc 1998: Lexical investments: French interlanguage as L2 versus L3. – *Applied Linguistics* 19 (4), s. 471–490.
- Dufva, Hannele, Aro, Mari, Suni, Minna & Salo, Olli-Pekka 2011: Onko kieltä olemassa? Teoreettinen kielitiede, soveltava kielitiede ja kielenoppimisen tutkimus. – *AFinLA-e Soveltavan kielitieteen tutkimuksia* n:o 3, s. 22–34.
- Ellis, Nick 2008: Phraseology: The periphery and the heart of language. – *Phraseology in foreign language learning and teaching* s. 1–13. Amsterdam: John Benjamins Publishing Company.
- Ellis, Rod 1985: *Understanding second language acquisition*. Oxford applied linguistics. Oxford: Oxford University Press.
- 1994: *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, Rod & Barkhuizen, Gary 2005: *Analysing learner language*. Oxford applied linguistics. Oxford: Oxford University Press.
- Faerch, Claus & Kasper, Gabriele 1984: Two ways of defining communication strategies. – *Language Learning* 34 (1) s. 45–63.
- Firth, John 1957: *Papers in linguistics 1934–1951*. London: Oxford University Press.
- 1968 [1957]: *A synopsis of linguistic theory, 1930–55*. – F. R. Palmer (ed.), *Selected papers of J. R. Firth 1952–59* s. 168–205. London: Longmans.
- Giddens, Anthony 1979: *Yhteiskuntateorian keskeisiä ongelmia. Toiminnan, rakenteen ja ristiriidan käsitteet yhteiskunta-analyysissä* (suom. Pasi Andersson ja Ilkka Heiskanen). Helsinki: Otava.
- Gilquin, Gaëtanelle & Paquot, Magali 2008: Too chatty: Learner academic writing and register variation. – *English Text Construction* 1 (1) s. 41–61. Draft, s. 1–26. – https://dial.uclouvain.be/pr/boreal/object/boreal%3A70006/datastream/PDF_01/view 28.5.2018.
- Granger, Sylviane 1997: On identifying the syntactic and discourse features of participle clauses in academic English: native and non-native writers compared. – Jan Aarts, Inge de Mönnink & Herman Wekker (eds.), *Studies in English language and teaching* s. 1–3. Amsterdam: Rodopi. Uudelleenjulkaisu verkossa. – <https://bit.ly/2NTNMX8> 30.7.2018.

- 1998: Prefabricated patterns in advanced EFL writing: Collocations and formulae. – Anthony Paul Cowie (ed.), *Phraseology: Theory, analysis and applications* s. 145–160. Oxford: Oxford University Press.
- 2002: A bird's eye view of learner corpus research. – Sylviane Granger, Heidrun Jung & Stephanie Petch-Tyson (eds.), *Computer learner corpora. second language acquisition and foreign language teaching* s. 3–33. Amsterdam: John Benjamins.
- 2005: Pushing back the limits of phraseology: How far can we go? – Chistelle Cosme, Celine Goucerneur, Fanny Meunier & Magali Paquot (eds.), *Pre-proceedings of phraseology 2005. The many faces of phraseology. An interdisciplinary conference. Louvain-la-Neuve Belgium, 13–15 October 2005* s. 165–168.
- Granger, Sylviane & Meunier, Fanny (toim.) 2008a: *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.
- 2008b: Phraseology in language learning and teaching. where to from here? – Fanny Meunier & Sylviane Granger (eds.), *Phraseology: An interdisciplinary perspective* s. 247–252. Amsterdam: John Benjamins
- 2015: Contrastive interlanguage analysis: a reappraisal. – *International Journal of Learner Corpus Research* 1 (1) s. 7–24.
- Granger, Sylviane & Paquot, Magali 2008: Disentangling the phraseological web. – Sylviane Granger & Fanny Meunier (eds.), *Phraseology. An interdisciplinary perspective* s. 27–49. Amsterdam: John Benjamins.
- Granger, Sylviane & Wynne, Martin 1998: Optimising measures of lexical variation in EFL learner corpora. – John Kirk (ed.), *Corpora galore. Analyses and techniques in describing English* s. 249–257. Amsterdam: Rodopi.
- Gries, Stefan 2008: Phraseology and linguistic theory. A brief survey. – Sylviane Granger & Fanny Meunier (eds.), *Phraseology. An interdisciplinary perspective* s. 3–36. Amsterdam: John Benjamins.
- Grönholm, Maija 1993: *Tv on pang pang – verbisanaston kehitys toisen kielen kirjoittamisessa*. Rapporter från Pedagogiska fakulteten vid Åbo Akademi 4. Vasa: Åbo Akademi.
- Hakulinen, Auli 2003: Vielä nykysuomesta ja sen huollosta. Ovatko puhuttu ja kirjoitettu kieli erkaantuneet toisistaan? – *Kielikello* (1). – <https://www.kielikello.fi/-/viela-nykysuomesta-ja-sen-huollosta-ovatko-puhuttu-ja-kirjoitettu-kieli-erkaantuneet-toisistaan> 26.5.2018.
- Hasselgren, Angela 1994: Lexical teddy bears and advanced learners: A study into the ways norwegian students cope with english vocabulary. – *International Journal of Applied Linguistics* 4 (2) s. 237–258.
- Heikkilä, Tarja 2005 [1998]: *Tilastollinen tutkimus*. 5.–6. painos. Helsinki: Edita.
- 2014: *SPSS. Riippuvuudet*. Tilastollinen tutkimus -kirjaan (2014, 9. painos) liittyvä verkkomateriaali. – www.tilastollinentutkimus.fi 28.7.2018.
- Heikkilä-Kopperoinen, Elina 2004: Virolainen suomenoppija ja demonstratiivin valinta. – Helena Sulkala & Heli Laanekask (toim.), *Virsu II: suomi ja viro kohdekielinä* s. 88–102. Lähivertailuja 15. Oulun yliopiston suomen ja saamen kielen ja logopedian laitoksen julkaisuja 24.
- Hoey, Michael 2005: *Lexical priming. A new theory of words and language*. London: Roudledge Falmer.
- Hoffmann, Sebastian Fischer-Starcke, Bettina & Sand, Andrea 2013: Introduction.. – *International Journal of Corpus Linguistics* 18 (1) s. 1–5.
- Howarth, Peter 1998: The phraseology of learners' academic writing. – Anthony Cowie (ed.), *Phraseology: Theory, analysis, and applications* s. 161–186. Oxford: Oxford University Press.
- Hunston, Susan 2002: *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hyland, Ken & Milton, John 1997: Qualification and certainty in L1 and L2 students' writing. – *Journal of Second Language Writing* 6 (2) s. 183–205.
- Ivaska, Ilmari 2015: *Edistyneen oppijansuomen konstruktiopiirteitä korpusvetoisesti: avainrakennanalyysi*. Väitöskirja. Turun yliopisto.
- Jantunen, Jarmo 2001: Tärkeä seikka ja keskeinen kysymys: Mitä korpuslingvistinen analyysi paljastaa lähisynonyymeista? – *Virittäjä* 105 (2) s. 170–192.
- 2004: *Synonymia ja käännessuomi: korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännessuomen leksikaalisiin erityispiirteisiin*. Väitöskirja. Joensuun yliopisto.
- 2007: Oppijansuomen piirteitä korpusvetoisesti. – Pirkko Muikku-Werner, Ossi Kokko & Hannu Remes (toim.), *Virsu III: suomalais-urkilaisia kohdekieliä ja kontakteja* s. 69–83. Studies in Languages 42. Joensuu: Joensuun yliopistopaino.
- 2008: Haasteita oppijankielen korpusanalyysille: oppijankielen universaalit. – Pille Eslon (toim.), *Öppijakeele analüüs: võimalused, probleemid, vajadused* s. 1–26. Tallinna Ülikooli eesti filoloogia osakonna toimetised 10. Tallinna: Tallinna Ülikool Kirjastus.
- 2009a: Ei pelkäästään mielikuvituksen puutteen vuoksi. – Kieliaineistojen systemaattinen käyttö kielentutkimuksessa. – *Virittäjä* 105 (2) s. 170–192.
- 2009b: "Minulla on aivan paljon rahaa": Fraseologiset yksiköt suomen kielen opetuksessa. – *Virittäjä* 113 (3) s. 356–381.
- 2011a: Avainsana-analyysi annotoidun oppijankieliaineiston tutkimisessa: Alustavia havaintoja. – *Sovelletun kielitieteen tutkimuksia. AFinLA-e* n:o 3 s. 48–61.

- 2011b: Kansainvälinen oppijansuomen korpus (ICLFI): typologia, taustamuuttujat ja annotointi. – *Lähi-vördlusi = Lähivertailuja* 21, s. 86–105.
- 2012: Akateemiset ja populaaritekstit korpusvertailussa: sanalistat, avainsanat ja fraseologiset yksiköt. – Vesa Heikkinen, Eero Voutilainen, Petri Lauerma, Ulla Tiirilä & Mikko Lounela (toim.), *Genreanalyysi – tekstilajitutkimuksen käytäntöä* s. 186–213. Kotimaisten kielten keskuksen verkkojulkaisuja 29. Helsinki: Kotimaisten kielten keskus.
- 2017: Lexical and morphological priming. A holistic phraseological analysis of the Finnish time expression kello. – Michael Pace-Sigge & Katie J. Patterson (eds.), *Lexical Priming: Applications and advances* s. 254–272. Amsterdam: John Benjamins.
- 2018: Homot ja heterot Suomi24:ssä: analyysi digitaalisista diskursseista. – *Puhe ja kieli* (1).
- Jantunen, Jarmo & Brunni, Sisko 2012: Morfologinen priming ja fraseologia vieraan kielen oppimisessa: Korpus-tutkimus oppijansuomesta. – *Lähi-vördlusi = Lähivertailuja* 22 s. 71–100.
- 2013: Morphology, lexical priming and second language acquisition: a corpus-study on learner Finnish. – Sylvia Granger, Gaëtanelle Gilquin & Fanny Meunier (eds.), *Twenty Years of Learner Corpus Research: Looking back, Moving ahead* s. 235–245. Louvain-la-Neuve: Presses universitaires de Louvain.
- Jantunen, Jarmo, Kumpulainen, Marjo, Tammimies, Tanja & Tokola, Teemu 2013: *Korpuspohjaista oppijansa-nakirjaa tekemässä: Esimerkkinä "ConLexis"*.
- Kaivapalu, Annekatrin 2005: *Lähdekieli kielenoppimisen apuna*. Jyväskylän yliopisto.
- Kangasniemi Heikki 1997: *Sana, merkitys ja maailma. Katsaus leksikaalisen semantiikan perusteisiin*. Helsinki: Finn Lectura.
- Karlsson, Fred 1994: *Yleinen kielitiede*. Helsinki: Yliopistopaino.
- 2006 [1994]: *Yleinen kielitiede*. Uudistetun laitoksen 3. painos. Helsinki: Yliopistopaino.
- Kennedy, Graeme 2008: Phraseology and language pedagogy: Semantic preference associated with english verbs in the british national corpus. – Fanny Meunier & Sylviane Granger (eds.), *Phraseology in foreign language learning and teaching* s. 21–41. Amsterdam: John Benjamins Publishing Company.
- Kenny, Dorothy 2001: *Lexis and creativity in translation. A corpus-based study*. Manchester: St. Jerome.
- Kielipankki 2004: Suomen sanomalehtikielen taajuussanasto. CSC – Tieteen tietotekniikan keskus. <https://korp.csc.fi/suomen-sanomalehtikielen-taajuussanasto-B9996.txt> 31.1.2018.
- Knight, Dawn, Evans, David, Carter, Ronald & Adolphs, Svenja 2009: HeadTalk, HandTalk and the corpus: towards a framework for multi-modal, multi-media corpus development. – *Corpora* 4 (1) s. 47–58. Amsterdam: Rodopi.
- Kortelainen, Kristiina 2011: *Ei ole kaikki muumit laaksossa: tutkimus suomen kielen idiomikonstruktion produktiivisuudesta*. Pro gradu -tutkielma. Turun yliopiston kieli- ja käännöstieteiden laitos.
- KS = *Kielitoimiston sanakirja*. Kotimaisten kielten keskuksen julkaisuja 166. Helsinki: Kotimaisten kielten keskus ja Kielikone 2018. – <https://www.kielitoimistonsanakirja.fi/> 15.2.2018.
- Kuiri, Kaija 2012: *Johdatus semantiikkaan*. Helsinki: Finn Lectura.
- Kuuluvainen, Helena 2015: *Fraseologiset virheet kansainvälisessä oppijansuomen korpuksessa*. Pro gradu -tutkielma. Oulun yliopisto.
- Larsen-Freeman, Diane 2014: Another step to be taken – Rethinking the end point of the interlanguage continuum. – ZhaoHong Han & Elaine Tarone (eds.), *Interlanguage. Forty years later* s. 203–220. Amsterdam: John Benjamins.
- Latomaa, Sirkku 1993: Mitä hyötyä on oppijoiden kielitaustan tuntemisesta? – Eija Aalto & Minna Suni (toim.), *Kohdekielenä suomi. Näkökulmia opetukseen* s. 9–31. Korkeakoulujen kielikeskusten selosteita N:o 1. Jyväskylä: Korkeakoulujen kielikeskus.
- 1996: Matkalla uuteen kieleen – Moneja baareja. Tiellä toimivaan kaksikielisyyteen. – Helena Ruuska & Sanna-Marja Tuomi (toim.), *Äidinkielen opettajain liiton vuosikirja XLII* s. 97–105.
- Latomaa, Sirkku & Tuomela, Veli 1993: Suomi toisena vai vieraana kielenä? – *Virittäjä* 97 (2) s. 238–245.
- Leech, Geoffrey 1992: Corpora and theories of linguistic performance. – Jan Startvik (ed.), *Directions in corpus linguistics: Proceedings of the nobel symposium 82, Stockholm, 4–8 August 1991* s. 105–122. Berlin: Mouton de Gruyter.
- 1998: Preface. – Sylvia Granger (ed.), *Learner English on computer* s. xiv–xx. Lontoo: Longman.
- Lorenz, Gunter 1999: Learning to cohere: causal links in native vs. non-native argumentative writing. – Wolfram Bublitz, Uta Lenk & Eija Ventola (eds.), *Coherence in spoken and written discourse*. Amsterdam: John Benjamins.
- Louw, Bill 1993: Irony in the text or insincerity in the writer? – The diagnostic potential of semantic prosodies. – Mona Baker, Gill Grancis & Elena Tognini-Bonelli (eds.), *Text and technology. In honour of John Sinclair* s. 157–176. Philadelphia: John Benjamins.
- Manninen, Terhi 2007: *Vieraskielisten ongelmia ylioppilasaineissa*. Pro gradu -tutkielma. Oulun yliopiston suomen ja saamen kielen ja logopedian laitos.

- Martin, Maisa & Alanen, Riikka 2011: Oppijankieli – kehitystä ja virheitä. – Paula Kajala, Riikka Alanen & Hannele Dufva (toim.), *Kieltä tutkimassa. tutkielman laatijan opas* s. 33–48. Helsinki: Finn Lectura.
- Mauranen, Anna 2000: Strange strings in translated language. A study on corpora. – Maeve Olohan (ed.), *Intercultural faultlines. Research models in translation studies I. Textual and cognitive aspects* s. 119–141. Manchester: St. Jerome.
- 2004: Spoken corpus for an ordinary learner. – John Sinclair (ed.), *How to use corpora in language teaching* s. 89–105. Amsterdam: John Benjamins.
- McEnery, Tony & Hardie, Andrew 2014 [2012]: *Corpus linguistics: Method, theory and practise*. 3. painos. Cambridge: Cambridge University Press.
- Mel'čuk, Igor 1995: The phrasemes in language and phraseology in linguistics. – Martin Everaert (ed.), *Idioms: Structural and psychological perspectives* s. 167–232. Hillsdale: Lawrence Erlbaum Associates.
- Meunier, Fanny & Granger, Sylviane 2008a (toim.): *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: Benjamins.
- Mikkonen, Miia 2007: ”Elämä ei oo tylsä”. *Oululaisten maahanmuuttajien puhekieli*. Pro gradu -tutkielma. Oulun yliopiston suomen kielen, informaatiotutkimuksen ja logopedian laitos.
- Mukherjee, Joybrato & Götz, Sandra 2015: Learner corpora and learning context. – Sylviane Granger, Gaëtanelle Gilquin & Fanny Meunier (eds.), *The Cambridge handbook of learner corpus research* s. 423–442. Cambridge: Cambridge University Press.
- Mustonen, Sanna 2015: *Käytössä kehittyvä kieli: Paikat ja tilat suomi toisena kielenä -oppijoiden teksteissä*. Väitöskirja. Jyväskylän yliopisto.
- Nesselhauf, Nadja 2003: The use of collocations by advanced learners of english and some implications for teaching. – *Applied Linguistics* 24 (2) s. 223–242.
- 2005: *Collocations in a learner corpus*. Studies in corpus linguistics. Amsterdam: Benjamins.
- Nieminen, Taija 2001: Kuvailun keinot oppijankielessä – merkitysten joustaminen ja intensiteetin ilmaiseminen espanjankielisten puhumassa suomessa. – Taija Nieminen (toim.) *Vuorovaikutus ja suomen kielen oppiminen* s. 76–126. Kakkoskieli 3. Helsingin yliopiston suomen kielen laitos.
- Nissilä, Leena 2011: *Viron kielen vaikutus suomen kielen verbien ja niiden rektioiden oppimiseen*. Väitöskirja. Oulun yliopisto.
- Ojutkangas, Krista, Larjavaara, Merja, Miestamo, Matti & Ylikoski, Jussi 2009: *Johdatus kielitieteeseen*. Helsinki: WSOY oppimateriaalit.
- Opetushallitus 2003: *Kielitaidon tasojen kuvausasteikko*. – http://www.edu.fi/download/119698_taitotasot.pdf 18.2.2018.
- 2018: *Ulkomaiset yliopistot, joissa voi opiskella suomea*. – <http://www.cimo.fi/ohjelmat/suomen-kieli-ja-kulttuuri/ulkomaiset-yliopistot-joissa-voi-opiskella-suomea> 31.1.2018.
- Palmer, Frank 1981: *Semantics. A New Outline*. 2. painos. Cambridge University Press.
- Paquot, Magali & Granger, Sylvia 2012: Formulaic language in learner corpora. – *Annual Review of Applied Linguistics* 32 s. 130–149.
- Partington, Alan 2004: ”Utterance content in each other’s company” – Semantic prosody and semantic preference. – *International Journal of Corpus Linguistics* 9 (1) s. 131–156.
- Pawley, Andrew & Syder, Frances 1983: Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. – Jack Richards & Richard Schmidt (eds.), *Language and communication* s. 191–225. London: Longman.
- Péry-Woodley, Marie-Paule 1990: Contrastive discourses: Contrastive analysis and a discourse approach to writing. – *Language Teaching* 23 s. 143–151.
- Pirkola, Silja 2016: *Synonyymien ehkä ja mahdollisesti kollokaatit ja semanttiset preferenssit*. Maisterintutkielma. Jyväskylän yliopisto.
- Seilonen, Marja 2011: *Olemattoman ilmaiseminen*. – Esitelmä Suomen kielen ja kulttuurin opettajien opinto- ja neuvottelupäivillä Kuopiossa 4.8.2011.
- Selinker, Larry 2014: Interlanguage 40 years on. Three themes from here. – Zhaohong Han & Elaine Tarone (eds.), *Interlanguage. Forty years later* s. 221–246, Amsterdam: John Benjamins.
- Seppälä, Tanja 2012: *Oppijansuomen kollokaatit ketjuuntuuissa verbirakenteissa*. Pro gradu -tutkielma. Oulun yliopisto.
- Sinclair, John 1987: *Looking up: an account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. London: Collins ELT.
- 1991: *Corpus, concordance, collocation. Describing English Language*. Oxford: Oxford University Press.
- 1996: The Search for Units of Meaning. – *Textus* IX s. 75–106.
- 2004: *Trust the text: Language, corpus and discourse*. London: Routledge.
- Spoelman, Marianne 2013: *Prior linguistic knowledge matters: The use of the partitive case in Finnish learner language*. Väitöskirja. Oulun yliopisto.
- Storhammar, Marja-Terttu 1994: Puhekielen asema ulkomaalaisopetuksessa. – *Virittäjä* 97 (1) s. 97–109.

- Strawn, Brent 2009: Comparative approaches: History, theory and the image of God. – Joel LeMon & Kent Richards (eds.), *Method Matters: Essays on the Interpretation of the Hebrew Bible in Honor of David L. Petersen* s. 117–142.
- Stubbs, Michael 1995: Collocations and semantic profiles: On the cause of the trouble with quantitative studies. – *Functions of Language* 2 (1) s. 2–24. Uudelleenjulkaistu verkossa. – <https://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1995-cause-trouble.pdf> 31.7.2018.
- 1996: *Text and Corpus Analysis. Corpus-assisted studies of language and culture*. Oxford: Blackwell.
- 2001: *Words and phrases. Corpus studies of lexical semantics*. Oxford: Blackwell Publishers.
- 2002: On text and corpus analysis: A reply to Borsley and Ingham. – *Lingua* 112 (1) s. 7–11.
- Tanner, Johanna 2008: S2-oppikirjojen dialogien tarkastelua. – Johanna Tanner & Marja Kokkonen (toim.), *Suomenoppiminen, kielitaito ja tutkimus* s. 163–189. Helsingin yliopisto: Suomen kielen ja kotimaisen kirjallisuuden laitos.
- Tarvainen, Jenny 2015: *Partitiiviobjektin semanttiset roolit: korpustutkimus*. Kandidaatintutkielma. Jyväskylän yliopisto.
- Tognini-Bonelli, Elena 1996: *Corpus theory and practise*. Birmingham: TWC.
- 2001: *Corpus linguistics at work*. Studies in corpus linguistics 6. Amsterdam: John Benjamins.
- 2008: *Phraseology as the starting point in a corpus-driven perspective*. Esitelmä konferenssissa International seminar on corpus linguistics: New trends in language teaching and translation studies. In honour of John Sinclair. Granada, Spain 22.–24. syyskuuta 2008.
- Tsui, Amy 2004: What teachers have always wanted to know – and how corpora can help. – John Sinclair (ed.), *How to use corpora in language teaching* s. 39–61. Amsterdam: John Benjamins.
- 2005: ESL teachers' questions and corpus evidence. – *International Journal of Corpus Linguistics* 10 (3) s. 335–356.
- Tieteen termipankki 2018: Kielitiede: syntagmaattinen suhde. – http://www.tieteentermipankki.fi/wiki/Kielitiede:syntagmaattinen_suhde 6.6.2018.
- 2018: Kielitiete: korpus. – <http://tieteentermipankki.fi/wiki/Kielitiede:korpus> 31.7.2018.
- Ullmann, Stephen 1967: *Semantics. An introduction to the science of meaning*. 3. painos. Oxford: Basil Blackwell.
- Valli, Raine 2001: *Johdatus tilastolliseen tutkimukseen*. 2. uudistettu painos. Jyväskylä: PS-kustannus.
- Vanhatalo, Ulla 2003: Kyselytestit vs. korpuslingvistiikka lähisynonyymien semanttisten sisältöjen arvioinnissa – mitä vielä keskeisestä ja tärkeästä? – *Virittäjä* 107 (3) s. 351–369.
- Varrio, Hanna 2014: *Sanoa-verbi fraseologisena yksikkönä oppijansuomessa ja natiivikielessä*. Pro gradu -tutkielma. Oulun yliopisto.
- Viinikka, Jenni, Voutilainen, Eero 2013: Ääniä ilmassa, merkkejä paperilla – puhutun ja kirjoitetun kielen suhteesta. – *Kielikello* (3). – <https://www.kielikello.fi/-/aania-ilmassa-merkkeja-paperilla-puhutun-ja-kirjoitetun-kielen-suhteesta> 26.5.2018
- VISK = Auli Hakulinen, Maria Vilkkunen, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen ja Irja Alho 2004: Iso suomen kielioppi. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkko-versio, viitattu 31.7.2018. Saatavissa: <http://scripta.kotus.fi/visk> URN:ISBN:978-952-5446-35-7.
- Wray, Alison 1999: Formulaic language in learners and native speakers. – *Language Teaching* 32 (4) s. 213–231.
- 2000: Formulaic sequences in second language teaching: Principle and practice. – *Applied Linguistics* 21 (4) s. 463–489.
- 2002: *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, Alison & Perkins, Michael R. 2000: The functions of formulaic language: an integrated model. – *Language & Communication* 20 (1) s. 1–28.
- Ädel, Annelie 2006: *The Use of Metadiscourse in Argumentative Texts by Advanced Learners and Native Speakers of English*. Amsterdam: John Benjamins.

Aineistot

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

Käännössuomen korpus. Käännössuomen sähköinen tutkimusaineisto. Käännössuomi ja kääntämisen universaalit -hankkeessa koostanut Joensuun yliopiston kansainvälisen viestinnänlaitos 1997-. Käyttöaika: [27.5.2018]. Saanti: Kansainvälisen viestinnän laitoksen palvelin, Joensuun yliopisto.

LIITE 1

Kielitoimiston sanakirjan (KS s.v. *saada*) kokonainen kuvaus verbistä SAADA.

saada (saamaton, saaminen, saatava, saatavana, saatavilla, saatavilta, saataville, saatavissa ks. erikseen)

1. tapahtumasta jossa jk esine, hyödyke, oikeus tms. siirtyy jkn t. jnk omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jk tulee. *Sai isältään kellon ja rahaa. Saada lahja, perintö. Ei saanut pankista lainaa. Saada hyvä virka, työpaikka. Saada jtak omakseen, käyttöönsä, lainaksi. Saada jtakperinnöksi (t. perintönä), lahjaksi (t. lahjana), palkinnoksi (t. palkintona). Ei saanut koko päivänä syödäkseen. Sai vastaanulijalta tulen tupakkaansa. Ovatko kaikkisaaneet kahvia? Saisinko [= antaisit(te)ko minulle] lasin vettä.*

Erik.

a. jnk lähetetyn vastaanottamisesta. *Saada kirje, viesti jkltä. Saada vastaus tiedusteluunsa. Kuitata lähetys saaduksi.*

b. jnk oikeuden, mahdollisuuden, etuisuuden tms. tulemisesta jkn osaksi. *Saada lupa, oikeus jhk. Sai kunnian toimia puheenjohtajana. Saada virkavapautta, lomaa. Sai eron vaimostaan. Pyysi ja sai puheenvuoron. Saat vielä yhden tilaisuuden yrittää. Saada hyvää palkkaa.*

2. tulla, joutua jnk toimenpiteen, käsittelyn, kohtelun tms. kohteeksi. *Saada syyte, haaste oikeuteen. Saada apua, hoitoa. Saada valmennusta, opetusta. Hyvänkasvatuksen saanut lapsi. Ehdotus sai innostuneen vastaanoton.*

Erik.

a. joutua kurituksen, pahoinpitelyn, moitteiden tms. kohteeksi. *Saada selkäsauna, selkäänsä. Saada keppiä t. kepistä, piiskaa t. piiskasta. Pojat saivat [kuritusta]oikein isän kädestä. Sai tytöltä ympäri korvia. Saada jkltä isku, potku. Saada nuhteita, moitteita, sapiskaa, kyytiä. Sai mitä ansaitsikin. Saada nenälleen t. nenilleentulla nolatuksi.*

Ark. *Saada turpiinsa, kuonoonsa, köniinsä. Siitäs sait, kun et uskonut. Saada siipeensä ks. siipi.*

b. tulla tuomitukseksi rangaistukseen. *Saada rangaistus, tuomio rikoksesta. Saada vankeutta, arestia, sakkoa. Sai neljä vuotta (vankeutta) petoksesta.*

c. tulla, joutua kiittävän, moittivan tms. suhtautumisen kohteeksi. *Saada osakseen kiitosta, myötätuntoa, halveksuntaa, pilkkaa. Mainetta saanut taiteilija.*

d. (koulu)arvosteluun t. arviointiin liittyen. *Saada hyvä todistus. Sai kokeessa nelosen. Saada ehdot. Mäkihyppääjä sai kaikilta tuomareilta 18 pistettä. Kirja sai hyväarvostelut.*

3. ilmaantua t. tulla jnk osaksi t. hyväksi. *Kuu saa valonsa auringosta. Saada väriä ihoonsa. Lapsi saa hampaita. Ravinnosta saatavat hivenaineet. Saada rohkeutta, voimaa jstak. Sai elämältä paljon. Saada uskonnollinen herätys. Sai päähänsä hullun idean. Saada harmia jstak. Sai surmansa kolarissa. Saatiin hyvät ilmat. Huhu saisiivet. Tapaus sai luonnollisen selityksen selittyi luonnolliseksi. Saada alkunsa alkaa; aiheutua, syntyä.*

Erik.

a. jälkeläisen syntymästä. *Nainen sai lapsen, pojan. Mies sai perillisen, lapsenlapsen. Sai lapsen tyttöystävänsä kanssa. Koira saa pentuja.*

b. sairauteen, oireeseen t. vammaan liittyen. *Saada tauti, kuumetta, päänsärkyä. Sai ruoasta mahanpuruja. Saada sydänkohtaus, halvaus. Sai lasinsirusta haavan.*

c. ark. hämmästyttä, raivostumista, pelästymistä, inhoamista tms. merkitsevissä ilmauksissa. *Saada kohtaus, halvaus, sätky, hepuli ks. kohtaus jne. Saan näppylöitä, kun näenkin hänet.*

d. muuttua, kehittyä jonkinlaiseksi. *Valukappale saa muotonsa muotissa. Uuden merkityksen saanut sana. Asia sai yllättävän käänteen.*

Mat. *Jos x :lle annetaan arvo 2, funktio saa arvon 0,5.*

4. onnistua hankkimaan jtkk itselleen, käyttöönsä, yhteyteensä tms., tavoittaa, saavuttaa. *Sai kirjan kymppillä, alennuksella. Sai halvalla hyvän auton. Odota, että ehdinsaa jotakin ylleni. Luki kaiken, minkä käsiinsä sai. Saada pakeneva kiinni. Sai kiinni köydestä. Saada jku valtaansa. Saada ote vastustajasta. Saada kosketus, yhteys, tuntuma jhk. Saada osuma maalitauluun. Juoksija sai hyvän lähdön onnistui lähdössä hyvin. Sai kultamitalin viimeisellä heitolla. Saada päävoitto arpajaisissa.*

Erik.

a. onnistua elatuksen, sadon, saaliin tms. hankkimisessa. *Saada toimeentulonsa kaupasta. Pellosta saatiin hyvä sato. Metsästäjä sai hyvän saaliin. Onki tunninsaamatta mitään.*

b. onnistua hankkimaan jku seurakseen t. ystävyys- tm. suhteeseen. *Saada jku toverikseen, saattajakseen. Hän on saanut paljon ystäviä, vihamiehiä kirjoituksillaan.*

c. onnistua hankkimaan jku puolisoon. *He saivat toisensa. Arka mies ei saa ko- reata akkaa SL.*

d. ark. päästä (naisen kanssa) yhdyntään. *Saitko sä siltä? Saada naista, miestä.*

e. ark. orgasmin saavuttamisesta. *Saada yhdynnässä. Saitko?*

5. jnk tulemisesta jkn aistimusten, tiedon t. ymmärryksen piiriin. *Sai pakenijan näkyviinsä. Saada korviinsa, tietoonsa ilkeitä huhuja. Saada tieto päätöksestä. Sai vihiäsuunnitelmasta. Saada käsitys, aavistus, vaikutelma, selko jstak. Ei saanut selvää käsialasta.*

6. tuloksen syntymisestä laskemalla, päättelämällä tms. tavalla. *Tutkimuksessa saatiin yllättäviä tuloksia. Kun kaksi kerrotaan kahdella, saadaan neljä. Yhtälöstä saadaan: n arvoksi viisi.*

7. jnk aseman, olotilan, tehtävän tms. tulemisesta jnk osaksi; saavuttaa, tavoittaa. *Muotokuva sai huoneessa kunniaapaikan. Maa sai itsenäisyyden. Tuli sai (yli)vallan. Saada jalansijaa. Ei saanut lapsilta hetkenkään rauhaa. Saada rauha sielulleen, tunnolleen. Saada anteeksi. Ei saanut unta ei kyennyt nukahtamaan. Saada jstakkyllikseen, kylläkseen, tarpeekseen kyllästyä jhk. Sai asian toimittakseen, toimitettavakseen. Sai muodostaakseen hallituksen. Toimekseen t. toimeksi saaneena N. N. [asiakirjan allekirjoituksena].*

8. onnistua, kyetä tekemään t. aiheuttamaan jtk t. saattamaan tekeminen päätökseen. *Korjaamo sai autosta vielä käyttökelpoisen. Luista saatiin hyvä keitto. Lapsia olivaikea saada vuoteeseen. Ei saanut asiaa mielestään. Etkö saa sanaa suustasi! Sai talonsa vesikattoon. Tehtyä ei saa tekemättömäksi. Saada jku hengiltä. Jännitys saihänet vapise- maan. Saada jtk aikaan. Sain vihdoon kirjeen aikaan Sinulle. Saada jtk aikaiseksi ark. aikaan. Saada jk esiin, irti, auki, rikki. Sai työn tehdyksi t. tehtyä. Vahingot saatiin korjatuksi t. korjattua. Sai kirjoituksensa julkaistuksi.*

9. ilmaisemassa että jk on luvallista, sallittua. *Saanko lainata kynääsi? Sai pitää löytämänsä esineen. Tili, jolta saa nostaa varoja rajoituksetta. Lämmityslaitetta ei saapeittää. Älä, ei saa lyödä!*

Erik. kohteliaisuutta ilmaisemassa. *Saanko tarjota lisää kahvia? Mitä Teille saisi olla [= mitä haluaisitte ostaa t. ottaa]? Saanen vielä korostaa yhtä seikkaa. Pyydänsaada huomauttaa, että --.*

10. ilmaisemassa, että jk sopii, ei tuota haittaa, on yhdentekevää. *Sato on korjattu, talvi saa tulla. Ellet usko, saat olla uskomatta. Nyt en saa epäonnistua.*

11. ilmaisemassa tilaisuutta, mahdollisuutta jhk; voida, saattaa. *Tilatut tavarat saa noutaa huomenna. Taikinan on saatava nousta kunnolla. Täällä saamme ollarauhassa. Ei saanut melulta t. melun vuoksi nukutuksi t. nukuttua.*

12. ilmaisemassa jnk pakollisuutta t. väistämättömyyttä; joutua tekemään jtak. *Tämän hän saa vielä maksaa. Sai tehdä paljon turhaa työtä. Sai nuolla näppejään. Tästä saat syyttää itseäsi. Saat uskoa, että kaikki vielä järjestyy. Aina häntä saa odottaa. Parempaa miestä saa hakea on vaikea löytää.*

13. ilmaisemassa (moitteeseen liittyvää) käskyä, kehotusta t. toivomusta. *Nyt saa riittää! Ellet tottele, saat jäädä ilman jälkiruokaa. Saat luvan pyytää anteeksi! Saisittepitää lapsenne paremmassa kurissa. Verotus saisi olla lievempää.*

14. nähdä-verbin ohella eräissä tulevaisuuteen viittaavissa ilmauksissa. *Saa nähdä* [= on jännittävää nähdä; on epävarmaa], *miten selviämme. Saatpa nähdä* [= huomaat vielä], *että olen oikeassa.* | *"Et onnistu."* *"Saadaanpa nähdä* [= se vielä nähdään]."

Erik. välittelevissä vastauksissa. *"Tuletko?" "Saa nyt nähdä."*

15. vanh., runok.

a. tulla, saapua. *Kun aamu saa.*

b. ryhtyä, ruveta. *Itse loihe lausumahan, sai itse sanelemahan KALEVALA.*

LIITE 2

Oppijansuomen aineiston SAADA-verbien merkitysluokat taajuusjärjestyksessä.

SIJA	POLYSEEMINEN LUOKKA	ABSOLUUTTINEN LUKUMÄÄRÄ	SUHTEELLINEN OSUUS
1.	1: tapahtumasta jossa jk esine, hyödyke, oikeus tms. siirtyy jkn t. jnk omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jk tulee.	144	28,1 %
2.	5: jnk tulemisesta jkn aistimusten, tiedon t. ymmärryksen piiriin.	86	16,8 %
3.	11: ilmaisemassa tilaisuutta, mahdollisuutta jhk; voida, saattaa.	58	11,3 %
4.	9: ilmaisemassa että jk on luvallista, sallittua.	45	8,8 %
5.	2: tulla, joutua jnk toimenpiteen, käsittelyn, kohtelun tms. kohteeksi.	42	8,2 %
6.	3: ilmaantua t. tulla jnk osaksi t. hyväksi.	37	7,2 %
7.	4: onnistua hankkimaan jtk itselleen, käyttöönsä, yhteyteensä tms., tavoittaa, saavuttaa	31	6,0 %
8.	8: onnistua, kyetä tekemään t. aiheuttamaan jtk t. saattamaan tekeminen päätökseen.	27	5,3 %
9.	7: jnk aseman, olotilan, tehtävän tms. tulemisesta jnk osaksi; saavuttaa, tavoittaa.	21	4,1 %
10.	10: ilmaisemassa, että jk sopii, ei tuota haittaa, on yhdentekevää.	10	1,9 %
11.	12: ilmaisemassa jnk pakollisuutta t. väistämättömyyttä; joutua tekemään jtk.	8	1,6 %
12.	6: tuloksen syntymisestä laskemalla, päättelemällä tms. tavalla.	2	0,4 %
13.	13: ilmaisemassa (moitteeseen liittyvää) käskyä, kehotusta t. toivomusta.	2	0,4 %
14.	14: nähdä-verbien ohella eräissä tulevaisuuteen viittaavissa ilmauksissa.	0	0,0 %
15.	15: vanh., runok.	0	0,0 %
	yht.	513	100,0 %

LIITE 3

Natiivisuomen aineiston SAADA-verbien merkitysluokat taajuusjärjestyksessä.

SIJA	POLYSEEMINEN LUOKKA	ABSOLUUTTINEN LUKUMÄÄRÄ	SUHTEELLINEN OSUUS
1.	8: onnistua, kyetä tekemään t. aiheuttamaan jtk t. saattamaan tekeminen päätökseen.	199	24,5 %
2.	1: tapahtumasta jossa jk esine, hyödyke, oikeus tms. siirtyy jkn t. jnk omistukseen, käyttöön, yhteyteen tms. ajateltuna sen kannalta, jonka omistukseen, käyttöön tms. jk tulee.	123	15,1 %
3.	3: ilmaantua t. tulla jnk osaksi t. hyväksi.	110	13,5 %
4.	4: onnistua hankkimaan jtk itselleen, käyttöönsä, yhteyteensä tms., tavoittaa, saavuttaa	66	8,1 %
5.	9. ilmaisemassa että jk on luvallista, sallittua.	60	7,4 %
6.	2: tulla, joutua jnk toimenpiteen, käsittelyn, kohtelun tms. kohteeksi.	56	6,9 %
7.	5. jnk tulemisesta jkn aistimusten, tiedon t. ymmärryksen piiriin.	56	6,9 %
8.	11. ilmaisemassa tilaisuutta, mahdollisuutta jhk; voida, saattaa.	52	6,4 %
9.	12. ilmaisemassa jnk pakollisuutta t. väistämättömyyttä; joutua tekemään jtk.	29	3,6 %
10.	7: jnk aseman, olotilan, tehtävän tms. tulemisesta jnk osaksi; saavuttaa, tavoittaa.	22	2,7 %
11.	6: tuloksen syntymisestä laskemalla, pääättelemällä tms. tavalla.	18	2,2 %
12.	10: ilmaisemassa, että jk sopii, ei tuota haittaa, on yhdentekevää.	11	1,4 %
13.	13: ilmaisemassa (moitteeseen liittyvää) käskyä, kehotusta t. toivomusta.	10	1,2 %
14.	14: nähdä-verbin ohella eräissä tulevaisuuteen viittaavissa ilmauksissa.	1	0,1 %
15.	15: vanh., runok.	0	0,0 %
	yht.	813	100,0 %

LIITE 4

SAADA-verbin merkitysluokkien suhteellisten osuuksien erojen tilastollinen merkitsevyys⁷¹ oppijan- ja natiivisuomessa. Oppijansuomen aineisto n = 513, natiivisuomen aineisto n = 813.

POLYSEEMINEN LUOKKA	Z-ARVO	KESKIVIRHE	P-ARVO	TILASTOLLINEN MERKITSEVYYS
8	-9,062435014	0,021201903	< 0,001	erittäin merkitsevä
1	5,723224135	0,022611423	< 0,001	erittäin merkitsevä
5	5,664212985	0,017435897	< 0,001	erittäin merkitsevä
3	-3,568752043	0,017702714	< 0,001	erittäin merkitsevä
11	3,157128184	0,015552042	< 0,01	merkitsevä
6	-2,654310495	0,006872439	< 0,01	merkitsevä
12	-2,161830189	0,00928649	< 0,05	melkein merkitsevä
13	-1,573416132	0,005339648	> 0,05	ei merkitsevä
4	-1,413425982	0,014682029	> 0,05	ei merkitsevä
7	1,389228656	0,009987846	> 0,05	ei merkitsevä
9	0,91415018	0,015225682	> 0,05	ei merkitsevä
2	0,88063225	0,014751511	> 0,05	ei merkitsevä
10	0,847088311	0,007039457	> 0,05	ei merkitsevä
14	-0,794652461	0,001547862	> 0,05	ei merkitsevä
15	-	-	-	-

⁷¹ Raja-arvot (z-arvolle): 1,96 melkein merkitsevä (p < 0,05); 2,57 merkitsevä (p < 0,01); 3,29 erittäin merkitsevä (p < 0,001). Positiivinen z-arvo kertoo, että luokka on suurempi oppijansuomen aineistossa ja negatiivinen z-arvo kertoo, että luokka on suurempi natiivisuomen aineistossa.

LIITE 5

Oppijansuomen morfologisen primingin jakautuminen taajuusjärjestyksessä.

SIJA	SANANMUOTO	ABSOLUUTTINEN LUKUMÄÄRÄ	SUHTEELLINEN OSUUS
1.	<i>saa</i>	28	19,4 %
2.	<i>saada</i>	28	19,4 %
3.	<i>saavat</i>	19	13,2 %
4.	<i>sai</i>	15	10,4 %
5.	<i>sain</i>	10	6,9 %
6.	<i>saanut</i>	6	4,2 %
7.	<i>saamme</i>	5	3,5 %
8.	<i>saan</i>	4	2,8 %
9.	<i>saisi</i>	4	2,8 %
10.	<i>saadakseen</i>	4	2,8 %
11.	<i>saaneet</i>	4	2,8 %
12.	<i>saimme</i>	3	2,1 %
13.	<i>saisin</i>	3	2,1 %
14.	<i>saadaan</i>	3	2,1 %
15.	<i>saisivat</i>	2	1,4 %
16.	<i>saamaan</i>	1	0,7 %
17.	<i>saamaansa</i>	1	0,7 %
18.	<i>saaman</i>	1	0,7 %
19.	<i>saamastani</i>	1	0,7 %
20.	<i>saatiin</i>	1	0,7 %
21.	<i>saivat</i>	1	0,7 %
	yht.	144	100,0 %

LIITE 6

Natiivisuomen morfologisen primingin jakautuminen taajuusjärjestyksessä.

SIJA	SANANMUOTO	ABSOLUUTTINEN LUKUMÄÄRÄ	SUHTEELLINEN OSUUS
1.	<i>saanut</i>	18	14,6 %
2.	<i>sai</i>	18	14,6 %
3.	<i>saada</i>	16	13,0 %
4.	<i>saat</i>	9	7,3 %
5.	<i>saa</i>	8	6,5 %
6.	<i>sain</i>	7	5,7 %
7.	<i>saivat</i>	7	5,7 %
8.	<i>saatiin</i>	5	4,1 %
9.	<i>saan</i>	4	3,3 %
10.	<i>saatu</i>	4	3,3 %
11.	<i>saavat</i>	4	3,3 %
12.	<i>saisi</i>	4	3,3 %
13.	<i>saadakseen</i>	3	2,4 %
14.	<i>saadulle</i>	1	0,8 %
15.	<i>saamaa</i>	1	0,8 %
16.	<i>saamaan</i>	1	0,8 %
17.	<i>saamaansa</i>	1	0,8 %
18.	<i>saamasta</i>	1	0,8 %
19.	<i>saaneensa</i>	1	0,8 %
20.	<i>saaneet</i>	1	0,8 %
21.	<i>saatava</i>	1	0,8 %
22.	<i>saatua</i>	1	0,8 %
23.	<i>saatuani</i>	1	0,8 %
24.	<i>saavansa</i>	1	0,8 %
25.	<i>saimme</i>	1	0,8 %
26.	<i>saisin</i>	1	0,8 %
27.	<i>saisinko</i>	1	0,8 %
28.	<i>saisivat</i>	1	0,8 %
29.	<i>sait</i>	1	0,8 %
	yht.	123	100,0 %

LIITE 7

Morfologisen primingin suhteellisten osuuksien erojen tilastollinen merkitsevyys⁷² oppijan- ja natiivisuomessa SAADA-verbin merkityksestä 1. Oppijansuomen aineisto n = 144, natiivisuomen aineisto n = 123.

SANANMUOTO	Z-ARVO	KESKIVIRHE	P-ARVO	TILASTOLLINEN MERKITSEVYYS
<i>saat</i>	-3,30214	0,022158567	< 0,001	erittäin merkitsevä
<i>saa</i>	3,085881	0,041934146	< 0,01	merkitsevä
<i>saanut</i>	-2,98073	0,035117155	< 0,01	merkitsevä
<i>saavat</i>	2,886173	0,034448427	< 0,01	merkitsevä
<i>saivat</i>	-2,3871	0,020931745	< 0,05	melkein merkitsevä
<i>saatu</i>	-2,1804	0,014914834	< 0,05	melkein merkitsevä
<i>saamme</i>	2,086226	0,01664356	< 0,05	melkein merkitsevä
<i>saatiin</i>	-1,85225	0,018197279	> 0,05	ei merkitsevä
<i>saadaan</i>	1,609851	0,012941159	> 0,05	ei merkitsevä
<i>saada</i>	1,413019	0,0455501	> 0,05	ei merkitsevä
<i>saaneet</i>	1,180498	0,01664356	> 0,05	ei merkitsevä
<i>saadulle</i>	-1,08404	0,007499829	> 0,05	ei merkitsevä
<i>saamaa</i>	-1,08404	0,007499829	> 0,05	ei merkitsevä
<i>saamasta</i>	-1,08404	0,007499829	> 0,05	ei merkitsevä
<i>saaneensa</i>	-1,08404	0,007499829	> 0,05	ei merkitsevä
<i>saatava</i>	-1,08404	0,007499829	> 0,05	ei merkitsevä
<i>saatua</i>	-1,08404	0,007499829	> 0,05	ei merkitsevä
<i>saatuani</i>	-1,08404	0,007499829	> 0,05	ei merkitsevä
<i>saavansa</i>	-1,08404	0,007499829	> 0,05	ei merkitsevä
<i>saisinko</i>	-1,08404	0,007499829	> 0,05	ei merkitsevä
<i>sait</i>	-1,08404	0,007499829	> 0,05	ei merkitsevä
<i>sai</i>	-1,0437	0,040408755	> 0,05	ei merkitsevä
<i>saaman</i>	0,925947	0,007499829	> 0,05	ei merkitsevä
<i>saamastani</i>	0,925947	0,007499829	> 0,05	ei merkitsevä
<i>saimme</i>	0,851719	0,014914834	> 0,05	ei merkitsevä
<i>saisin</i>	0,851719	0,014914834	> 0,05	ei merkitsevä
<i>saisivat</i>	0,444999	0,012941159	> 0,05	ei merkitsevä
<i>sain</i>	0,4181	0,029978164	> 0,05	ei merkitsevä
<i>saan</i>	-0,22657	0,020931745	> 0,05	ei merkitsevä
<i>saisi</i>	-0,22657	0,020931745	> 0,05	ei merkitsevä
<i>saadakseen</i>	0,172678	0,019617617	> 0,05	ei merkitsevä
<i>saamaan</i>	-0,112	0,010586405	> 0,05	ei merkitsevä
<i>saamaansa</i>	-0,112	0,010586405	> 0,05	ei merkitsevä

⁷² Raja-arvot (z-arvolle): 1,96 melkein merkitsevä (p < 0,05); 2,57 merkitsevä (p < 0,01); 3,29 erittäin merkitsevä (p < 0,001). Positiivinen z-arvo kertoo, että luokka on suurempi oppijansuomen aineistossa ja negatiivinen z-arvo kertoo, että luokka on suurempi natiivisuomen aineistossa.

LIITE 8

Sanojen aakkostettu luokittelu semanttisiin ryhmiin semanttisen primingin analyysia varten.

SEMANTTINEN LUOKKA	ICLFI	KSK
'abstrakti entiteetti'	asioita (x2), matkustamiseen, nimensä (x3), palvelua, palveluihin (x2), sukunimiä, tapahtumien, tehtävän, todisteita, urheilun	etunimensä, ihotautien, kyydin, lupa, luvan, matka, matkansa, nimensä, tilaisuutta, todistetta
'additiivisuus'	eli, ja (x15), jne., lisäksi, myös (x5), niinkin, sekä (x2), vielä	ja (x14), kaiken kaikkiaan, kanssa, lisäksi, mukaansa, muustakin, muutakin, toiset, vielä, yhteiseksi
'aika, ikä'	1970-luvulla, aina, ennen, jatkossa, joka kerta, joka kuukausi (x2), joskus (x2), jouluna, jälkeen, kauan sitten, ketkeksi (*hetkeksi), kun, lopulta, lähitulevaisuudessa, nyt (x2), pian, päivällä, päivänä (x2), siinä ajassa (*siinä ajassa), silloin (x3), sitten (x4), syntymäpäivä, syntymäpäiväksi, taas, tavallisesti, teini-ikäisenä, toisinaan, tänään, usein (x2), uuden (x2), vanhin, vielä, viikonloppuna, vuodelta, vuotta (x2), yhtäkkiä	1558, 1780, 1996, aikaa, aikana, aikoihin, ehdin, eilän (*eilen), ensimmäiseltä, ensin (x2), enää, illalla, jouluaattona (x2), juuri, kerralla, kohta, kuluttua, kun, lopullisia, mennessä, nyt (x2), päätteeksi, resptiittiäikää, samalla, seuraavana päivänä, sitten, taas, tänä päivänä, usein, uudelleen (x2), uuden (x2), uutta, vanhan, vastaisuudessakin, vuoden, vuonna (x2), vuoroa
'ihmiset'	Annikin, Anssille, F. E. Sillanpää, he (x4), hotajaa (*hoitajaa), hän (x15), hänelle, hänen (x3), ihminen, ihmisen, ihmiset, ihmisiä, itselleen, Jannaa, jokainen (x2), jotkut, joulupukista, kaiket, Kaisan (x2), kukin, maahanmuuttajan, Maiju Lassilan, miehet, mies, Minna Canthin, minulta, minä (x2), minähän, naiset (x2), olen, osallistujat, sankarit, sinkku, sinulta, suomalaiset (x2), Timo (x2), venäläiset, venäläisiltä, Vladimir Putin (x2)	aikuiset, Asko, don Salvatore, ei kukaan, emännältä, Esku, Hannibal, he (x2), henkilöitä, Hilarioiden, hoitaja, hän (x4), häneltä (x2), ihmisessä, jokainen, joku, Koskinen, kummisetä, Lomakka, matkailija, meiltä, mia (*minä), minulta, minun, minä (x2), minäkin, olin, Pau-liina, Petteri, rengin, sinä (x3), Sluppi, Syysprinssi, takavarikoijat, uhrit, Varjokuu

'kausalisuus'	jotta, koska (x3), kun (x2), siis, siksi, sitten	– (sitten, takia → 'luokittelemattomat')
'kommunikaatio'	esitellä, esittävä, kirjeen (x3), kirjeitä, kutsuneet, puhelinnumero, puhelinnumeron, puhua, sanaa, sanottiin, suomea (x2), uutisen, vastauksen	edustaa, kirjeen, kirjeitä, kutsuja, kuvanvaihdossa, luettavaa, murteella, postia, sanoi, sanoman, suunnuoron, vastaukseksi (x4), vastauksen, vastauksia
'konkreettinen entiteetti'	arpalippua, avaimet, elokuva, kalan, koiran, koiranpennun, kolmion, kukkia, meripihkaa, sähköä, tavaroita	ajokoirakin, ATK-järjestelmänsä, avainta, fluoria, hiireltä, huuliharpun, kasetin, lasivitrinikaappi, lorukepin, maata, mandoliinin, narunpätkä, paketin, partioveneensä, pe-toa, poroja, rasian, saippuaa, sähköä, tuolinsa, varastojen, välineistäni, öljyä
'koulutus'	lukion, opiskelija (x4), opiskella, yliopistossa	–
'laatu, ominaisuus'	harvinaisiin, helposti, hyviä (x2), hyvän, kiva (x2), omaperäisiä, parempaa (x3), parempia, samanlaisen, sellainen, vaikea (x2)	harvinaisen, kelvollisia, kirkkaimman, kylmän, matalasti, meluavamman, turhanpäiväisestä, turvallisempi, välittömistä, yksilöllisen
'lahja, palkinto'	bonuspisteitä, joululahja, joululahjoja, Jussi-palkinnot, kansallisstipendin (*kansallisstipendin) (x2), lahjaksi (x4), lahjan, lahjat, lahjoja (x4), Nobelin kirjallisuuspalkinnon, perintönä	kruunun, palkinto, palkintojakin, perintönä
'mentalisuus'	haluaa (x2), haluaisin, haluamansa, haluan, haluat, halusin, ilahduttava, mielestäni, onnellinen, vaikea, valitsisin, vihainen	erehdyttyään, mainetta, myöntävän, toivomuksen, turhaan, vaivalla, yllättäen, yrittää
'modalisuus'	ehkä, pitäisi, pitää, tietysti, toivottavasti, täytyy, voi, voidaan (x2), voisi (x2), voisin (x2)	jos (x4), kuulemma, pitäisi (x3), pitää, saattoi, taisi, tulisi, täytyisi, varmasti (x2), voisin, voitu
'määrä, intensiteetti'	aika, aika paljon, enemmän (x4), kaksi (x2), kokoon (*kokoon), lasin, liikaa (x3), miljoonien, molemmat, muutamien, niin, oikein, osa, paljon (x2), pienempää, pienehen, prosenttia, suurin, tarpeeksi,	2, 2,5, 30, hieman, ihan, joukon, kaiken, kaksi, koko, kokoon, kyllä, lästin, miljoonaa, osansa, osuutensa, parttion, pienempi, pienen (x2), suurempi, suuren, suuria, suuruinen, vain, vähemmän, vähän

	toisen (x2), vain, viisi (x2), yhden, yhä enemmän, yksitoista	
'negaatio'	ei (x5), ei edes, ei ole, eivät (x2), en, en ole, et (x3), ettei, kuitenkin (x2), mutta (x6), paitsi, päinvastoin	ei (x3), ei... ole, en, en ole, joka tapauksessa, mutta (x3), tai, vaan (x2)
'oikeus, mahdollisuus'	mahdollista (x2), mahdollisuuden (x2), mahdollisuus (x2), oikeus, pääsyä, tilaisuuden, tilaisuus	– (lupa, tilaisuutta → 'abstrakti entiteetti')
'paikka, suunta'	Bulgariassa, Etelä-Afrikassa, Helsingistä, Hovimäki, kierrätyskeskuksesta, Kimberlyessä, kotonan, mistä tahansa, Norjassa, päälle, päärakennuksista, siellä (x2), siellähän, sinne, Suomessa, Suomesta, Suomi, Yhdysvalloissa	apteekkiin, Bronxin-päämajaansa, eteensä, Hervantaan, Kapakassakin, kaupunki, Kotkan, kotoa, kovimman kautta, päällä, ratsastusleirille, Ratsutila, sahateollisuuskaupunki, siellä, sinne, Suomesta, takaisin, Turun (x2), Vaasan, Valamon
'perhe'	lapset, perheensä, poika, vaari, vanhemmat, vanhemmatkin	isä, lapset, vanhemmilleen, vanhempien, äidin
'prosessi'	alkoi, käytäisimme, saavat, yrittänyt, yrittävät	aloitan, hoitoonsa, nouseva, remontista, tappelun, toimittuaan, tulla, täydennystä, urheillut, yritti
'raha'	bafögia (x3), jatkorajoitushakemuksia, kultaa, laina, lapsilisän, opintolainan, opintorahaa, osataakseen, palkkaa (x3), palkkiota, palkkoja, rahaa (x12), rahallisesti, rahallista, rahoitusta, summa, tuen, veroista (x2)	apurahoja, lisätuloina, palkan, palkkaa, rahaa (x2), rahapalkkaa, remonttirahat, taalarin, tullitaksan, tuontitullitaksan, veroista
'relaationaalisuus'	joka (x2), jonka, jossa, jota, jotka (x4), mitä (x5), niistä, niitä, se (x2), sen, sille, tämän (x2)	joissa, joka (x3), jokin, jolla, jonka (x3), jossa, jostakin, jotka, ne, näiden, se (x2), sen (x3), siihenkin, siitä (x4), tämän, tästä
'ruoka'	ruokaa, suklaamunat, alkoholi, lauantaimekkaraa, maitoa, makkaransa, perunaa, kahvikuppi, sampanjaa	bouillabaisse-keittoa, mahani, pizzan, syödäkseni, söi, vanikkaa, äidinmaidon korviketta
'tila'	oli, olosuhteita, on (x4), paenneelta, säilyä, tilaa, vapautta (x2)	elämistä, kaadetun, kuollut, lomaa, olevat, oma, oman (x3), on, valvontaansa, vapauden, voi

'työ'	ammatin, asiantuntijat, työ, työkokemusta, työn (x2), työntekijät, työskentele, työtä (x3), viran (x2; virkan *viran), ylennystäkään	ammatin, kokoukseen, tullikirjuri, työstään, virkaa
'yhteiskunnallisuus, instituutio-naalisuus'	kansalle, pankilta, pääministeri, valtiolta (x3), valtionjohtajan, valtiosta, viranomaisille, yhteiskunta (x2)	armeija, kansainvälisessä, kaupasta, kirkkoon, näytösoikeudenkäynnit, ortodoksiseen, perinteestä, sosiaaliturvaa, Suomi (x2), tapulioikeutensa, tullikamarit, tullilaboratorio, viisumin
'luokittelemattomat'	että (x8), olen, oli, olivat (x2), on (x2), ovat (x3), vaikkapa	että (x5), kuin, kuten, oli (x9), olisikin, ollut, on, sitten, takia

LIITE 9

Oppijansuomen semanttisen primingin jakautuminen taajuusjärjestyksessä.

SIJA	SEMANTTINEN LUOKKA	ABSOLUUTTINEN LUKUMÄÄRÄ	SUHTEELLINEN OSUUS
1.	'ihmiset'	66	13,9 %
2.	'aika, ikä'	47	9,9 %
3.	'määrä, intensiteetti'	34	7,2 %
4.	'raha'	34	7,2 %
5.	'additiivisuus'	27	5,7 %
6.	'negaatio'	25	5,3 %
7.	'relaationaalisuus'	22	4,6 %
8.	'paikka, suunta'	19	4,0 %
9.	'lahja, palkinto'	18	3,8 %
10.	'kommunikaatio'	16	3,4 %
11.	'laatu, ominaisuus'	16	3,4 %
12.	'abstrakti entiteetti'	14	2,9 %
13.	'työ'	14	2,9 %
14.	'mentaalisuus'	13	2,7 %
15.	'modaalisuus'	13	2,7 %
16.	'konkreettinen entiteetti'	11	2,3 %
17.	'tila'	11	2,3 %
18.	'yhteiskunnallisuus, institutionaalisuus'	11	2,3 %
19.	'oikeus, mahdollisuus'	10	2,1 %
20.	'kausaalisuus'	9	1,9 %
21.	'ruoka'	9	1,9 %
22.	'koulutus'	7	1,5 %
23.	'perhe'	6	1,3 %
24.	'prosessi'	5	1,1 %
25.	'luokittelemattomat'	18	3,8 %
	yht.	475	100,0 %

LIITE 10

Natiivisuomen semanttisen primingin jakautuminen taajuusjärjestyksessä.

SIJA	SEMANTTINEN LUOKKA	ABSOLUUTTINEN LUKUMÄÄRÄ	SUHTEELLINEN OSUUS
1.	'ihmiset'	44	11,8 %
2.	'aika, ikä'	44	11,8 %
3.	'määrä, intensiteetti'	26	7,0 %
4.	'relaationaalisuus'	26	7,0 %
5.	'additiivisuus'	23	6,2 %
6.	'konkreettinen entiteetti'	23	6,2 %
7.	'paikka, suunta'	21	5,6 %
8.	'kommunikaatio'	17	4,6 %
9.	'modaalisuus'	17	4,6 %
10.	'yhteiskunnallisuus, institutionaalisuus'	14	3,8 %
11.	'negaatio'	13	3,5 %
12.	'tila'	13	3,5 %
13.	'raha'	12	3,2 %
14.	'abstrakti entiteetti'	10	2,7 %
15.	'laatu, ominaisuus'	10	2,7 %
16.	'prosessi'	10	2,7 %
17.	'mentaalisuus'	8	2,1 %
18.	'ruoka'	7	1,9 %
19.	'perhe'	5	1,3 %
20.	'työ'	5	1,3 %
21.	'lahja, palkinto'	4	1,1 %
22.	'luokittelemattomat'	21	5,6 %
	yht.	373	100,0 %

LIITE 11

Semanttisen primingin suhteellisten osuuksien erojen tilastollinen merkitsevyys⁷³ oppijan- ja natiivisuomessa SAADA-verbin merkityksestä 1. Oppijansuomen aineisto n = 475, natiivisuomen aineisto n = 373.

SEMANTTINEN LUOKKA	Z-ARVO	KESKI-VIRHE	P-ARVO	TILASTOLLINEN MERKITSEVYYS
'konkreettinen entiteetti'	-2,836983171	0,013572271	< 0,01	merkitsevä
'oikeus, mahdollisuus'	2,818925043	0,007468319	< 0,01	merkitsevä
'kausaalisuus'	2,672672909	0,007089296	< 0,01	merkitsevä
'raha'	2,514840524	0,015669926	< 0,05	melkein merkitsevä
'lahja, palkinto'	2,470594255	0,010997709	< 0,05	melkein merkitsevä
'koulutus'	2,354271576	0,006259619	< 0,05	melkein merkitsevä
'prosessi'	-1,785559771	0,009109969	> 0,05	ei merkitsevä
'työ'	1,569387202	0,010238938	> 0,05	ei merkitsevä
'relaationaalisuus'	-1,463021943	0,015986981	> 0,05	ei merkitsevä
'modaalisuus'	-1,424701197	0,012780214	> 0,05	ei merkitsevä
'luokittelemattomat'	-1,270108843	0,014491302	> 0,05	ei merkitsevä
'negaatio'	1,242146876	0,014313148	> 0,05	ei merkitsevä
'yhteiskunnallisuus'	-1,228445032	0,011702288	> 0,05	ei merkitsevä
'paikka, suunta'	-1,111368846	0,014666839	> 0,05	ei merkitsevä
'tila'	-1,019335688	0,011472817	> 0,05	ei merkitsevä
'ihmiset'	0,902779248	0,023244777	> 0,05	ei merkitsevä
'kommunikaatio'	-0,88884384	0,0133794	> 0,05	ei merkitsevä
'aika, ikä'	-0,88803295	0,021412604	> 0,05	ei merkitsevä
'laatu, ominaisuus'	0,576396601	0,011926786	> 0,05	ei merkitsevä
'mentaalisuus'	0,550693298	0,010751356	> 0,05	ei merkitsevä
'abstrakti entiteetti'	0,475615273	0,01123807	> 0,05	ei merkitsevä
'additiivisuus'	-0,295779539	0,016296236	> 0,05	ei merkitsevä
'määrä, intensiteetti'	0,105632134	0,017739427	> 0,05	ei merkitsevä
'perhe'	-0,098777539	0,007828164	> 0,05	ei merkitsevä
'ruoka'	0,019187793	0,00941288	> 0,05	ei merkitsevä

⁷³ Raja-arvot (z-arvolle): 1,96 melkein merkitsevä (p < 0,05); 2,57 merkitsevä (p < 0,01); 3,29 erittäin merkitsevä (p < 0,001). Positiivinen z-arvo kertoo, että luokka on suurempi oppijansuomen aineistossa ja negatiivinen z-arvo kertoo, että luokka on suurempi natiivisuomen aineistossa.

LIITE 12

Oppijansuomen morfologisen ja semanttisen primingin X^2 -riippumattomuustesti.

Ristiintaulukointi, odotusarvot ja kontribuutioarvot:

MP/SP	1 ABST.	2 KONKR.	3 IHMISET + PERHE	4 KIELIOPILLIS- SEMANTTISET	5 MÄÄRÄ + ADDITIIVISUUS	6 AIKA	7 MENTAALISUUS + KOMMUNIK.	8 PAIKKA, SUUNTA + TILA	9 MUUT	summa
1 SAA, SAI	30	28	27	19	22	13	8	8	8	163
2 SAADA	17	11	9	20	10	8	9	13	5	102
3 SAAVAT	7	13	22	9	12	6	2	2	2	75
4 SAIN, SAAMME, SAAN	5	10	4	7	13	15	5	2	4	65
5 SAANUT, SAANEET, SAISI, SAADAKSEEN	13	10	10	5	4	5	5	5	13	70
summa	72	72	72	60	61	47	29	30	32	475
Expected	1 ABST.	2 KONKR.	3 IHMISET + PERHE	4 KIELIOPILLIS- SEMANTTISET	5 MÄÄRÄ + ADDITIIVISUUS	6 AIKA	7 MENTAALISUUS + KOMMUNIK.	8 PAIKKA, SUUNTA + TILA	9 MUUT	summa
1 SAA, SAI	24,70737	24,707368	24,70736842	20,58947368	20,93263158	16,1284	9,951578947	10,29473684	10,9810526	163
2 SAADA	15,46105	15,461053	15,46105263	12,88421053	13,09894737	10,0926	6,227368421	6,442105263	6,87157895	102
3 SAAVAT	11,36842	11,368421	11,36842105	9,473684211	9,631578947	7,42105	4,578947368	4,736842105	5,05263158	75
4 SAIN, SAAMME, SAAN	9,852632	9,8526316	9,852631579	8,210526316	8,347368421	6,43158	3,968421053	4,105263158	4,37894737	65
5 SAANUT, SAANEET, SAISI, SAADAKSEEN	10,61053	10,610526	10,61052632	8,842105263	8,989473684	6,92632	4,273684211	4,421052632	4,71578947	70
summa	72	72	72	60	61	47	29	30	32	475
Contributions	1 ABST.	2 KONKR.	3 IHMISET + PERHE	4 KIELIOPILLIS- SEMANTTISET	5 MÄÄRÄ + ADDITIIVISUUS	6 AIKA	7 MENTAALISUUS + KOMMUNIK.	8 PAIKKA, SUUNTA + TILA	9 MUUT	summa
1 SAA, SAI	1,133749	0,438793	0,212737	0,122705	0,054426	0,606818	0,382719	0,511506	0,809273	4,272726
2 SAADA	0,153182	1,287169	2,700023	3,929962	0,733149	0,433891	1,234468	6,675765	0,509753	17,65736
3 SAAVAT	1,678606	0,234162	9,942495	0,023684	0,582399	0,272116	1,452511	1,581287	1,844298	17,61156
4 SAIN, SAAMME, SAAN	2,390025	0,002204	3,476563	0,178475	2,593270	#####	0,268156	1,079622	0,032794	21,43632
5 SAANUT, SAANEET, SAISI, SAADAKSEEN	0,538106	0,035129	0,035129	1,669486	2,769333	0,535738	0,123438	0,075815	14,552843	20,33502
summa	5,893668	1,9974578	16,36694756	5,924312388	6,732576245	13,2638	3,461291311	9,92399437	17,7489613	81,31299

Riippumattomuusarvot:

X ²	81,31298534706600000
p	0,00000358667489556
df	32,00000000000000000
X ² at 5%	46,19425952027850000

Luokkien selitteet:

sanamuodot

1. *saa + sai*
2. *saada*
3. *saavat*
4. *sain + saamme + saan*
5. *saanut + saaneet + saisi + saadakseen*

semanttiset luokat

1. 'abstrakti entiteetti' + 'yhteiskunnallisuus, institutionaalisuus' + 'työ' + 'koulutus' + 'laatu, ominaisuus' + 'oikeus, mahdollisuus'
2. 'konkreettinen entiteetti' + 'lahja, palkinto' + 'raha' + 'ruoka'
3. 'ihmiset' + 'perhe'
4. 'negaatio' + 'relaationaalisuus' + 'modaalisuus'
5. 'määrä, intensiteetti' + 'additiivisuus'
6. 'aika, ikä'
7. 'mentaalisuus' + 'kommunikaatio'
8. 'paikka, suunta' + 'tila'
9. 'luokittelemattomat' + 'prosessi' + 'kausaalisuus'

LIITE 13

Natiivisuomen morfologisen ja semanttisen primingin X^2 -riippumattomuustesti.

Ristiintaulukointi, odotusarvot ja kontribuutioarvot:

MP/SP	1 AIKA + PAIKKA, TILA	2 ABST.	3 KIELIOPILLISET	4 MÄÄRÄ + ADD.	5 IHMISET + PERHE	6 KONKR.	7 MUUT	summa	
1 SAI + SAA		27	20	9	11	13	13	5	98
2 SAANUT		7	12	9	5	11	10	14	68
3 SAAT + SAATIIN + SAISI + SAATU		10	7	14	12	6	12	4	65
4 SAADA		16	8	15	7	6	9	3	64
5 SAIVAT + SAAVAT		9	6	5	10	8	3	2	43
6 SAIN + SAAN		10	6	4	3	5	4	3	35
summa		79	59	56	48	49	51	31	373
Expected	1 AIKA + PAIKKA, TILA	2 ABST.	3 KIELIOPILLISET	4 MÄÄRÄ + ADD.	5 IHMISET + PERHE	6 KONKR.	7 MUUT	summa	
1 SAI + SAA	20,75603217	15,50134048	14,71313673	12,61126005	12,8739946	13,3994638	8,14477212	98	
2 SAANUT	14,40214477	10,75603217	10,20911528	8,750670241	8,93297587	9,29758713	5,65147453	68	
3 SAAT + SAATIIN + SAISI + SAATU	13,76675603	10,28150134	9,758713137	8,36461126	8,53887399	8,88739946	5,40214477	65	
4 SAADA	13,55495979	10,1233244	9,608579088	8,235924933	8,4075067	8,75067024	5,31903485	64	
5 SAIVAT + SAAVAT	9,107238606	6,801608579	6,455764075	5,533512064	5,64879357	5,87935657	3,57372654	43	
6 SAIN + SAAN	7,412868633	5,536193029	5,254691689	4,504021448	4,59785523	4,78552279	2,90884718	35	
summa	79	59	56	48	49	51	31	373	
Contributions	1 AIKA + PAIKKA, TILA	2 ABST.	3 KIELIOPILLISET	4 MÄÄRÄ + ADD.	5 IHMISET + PERHE	6 KONKR.	7 MUUT	summa	
1 SAI + SAA	1,878352	1,305560	2,218421	0,205860	0,001233	0,011909	1,214226	6,835562	
2 SAANUT	3,804416	0,143869	0,143201	1,607594	0,478294	0,053066	12,332689	18,56313	
3 SAAT + SAATIIN + SAISI + SAATU	1,030631	1,047342	1,843329	1,579996	0,754887	1,090114	0,363931	7,71023	
4 SAADA	0,441036	0,445358	3,025153	0,185469	0,689394	0,007104	1,011071	5,804586	
5 SAIVAT + SAAVAT	0,001263	0,094474	0,328272	3,605217	0,978646	1,410136	0,693006	7,111016	
6 SAIN + SAAN	0,902923	0,038856	0,299590	0,502236	0,035173	0,128940	0,002856	1,910574	
summa	7,15569767	3,03660388	7,558376053	7,184137225	2,90245474	2,57232936	15,6149235	46,02452	

Riippumattomuusarvot:

X ²	46,02452245921950000
p	0,03090738486053680
df	30,00000000000000000
X ² at 5%	43,77297182574220000

Luokkien selitteet:

sananmuodot	semanttiset luokat
1. <i>sai + saa</i>	1. 'aika, ikä' + 'paikka, suunta' + 'tila'
2. <i>saanut</i>	2. 'abstrakti entiteetti' + 'kommunikaatio' + 'mentaalisuus' + 'työ' + 'laatu, ominaisuus' + 'yhteiskunnallisuus, institutionaalisuus'
3. <i>saat + saatiin + saisi + saatu</i>	3. 'negaatio' + 'relaationaalisuus' + 'modaalisuus'
4. <i>saada</i>	4. 'määrä, intensiteetti' + 'additiivisuus'
5. <i>saivat + saavat</i>	5. 'ihmiset' + 'perhe'
6. <i>sain + saan</i>	6. 'konkreettinen entiteetti' + 'lahja, palkinto' + 'raha' + 'ruoka'
	7. 'luokittelemattomat' + 'prosessi'

