

AAMIAISKAHVILASTA ÖTÖKKÄTARJONTAAN

Suomen kirjoitetun yleiskielen morfosyntaktisten
yhdyssanarakenteiden produktiivisuus



Turun yliopisto
University of Turku

AAMIAISKAHVILASTA ÖTÖKKÄTARJONTAAN

Suomen kirjoitetun yleiskielen morfosyntaktisten
yhdyssanarakenteiden produktiivisuus

Laura Tyysteri

Turun yliopisto

Humanistinen tiedekunta

Kieli- ja käännöstieteiden laitos

Suomen kielen ja suomalais-ugrilaisen kielentutkimuksen oppiaine

Tohtoriohjelma Utuling

Työn ohjaajat:

Professori Kaisa Häkkinen,
Turun yliopisto

Universitetslärare Tomas Lehecka,
Åbo Akademi

Tarkastajat:

Professori Klaus Laalo,
Tampereen yliopisto

Professori Jarmo Jantunen,
Jyväskylän yliopisto

Vastaväittäjä:

Professori Klaus Laalo,
Tampereen yliopisto

Kustos:

Professori Kaisa Häkkinen
Turun yliopisto

© 2015 Laura Tyysteri

Taitto: Eeva Herrala

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck-järjestelmällä.

ISBN 978-951-29-6183-2 (PRINT)

ISBN 978-951-29-6184-9 (PDF)

ISSN 0082-6995

Painosalama Oy – Turku, 2015

TURUN YLIOPISTO

Humanistinen tiedekunta

Kieli- ja käännöstieteiden laitos

Suomen kieli ja suomalais-ugrilainen kielentutkimus

TYYSSTERI, LAURA: *Aamiaiskahvilasta ötökkätarjontaan*. Suomen kirjoitetun yleiskielen morfosyntaktisten yhdyssanarakenteiden produktiivisuus

Väitöskirja, 242 s.

Tohtoriohjelma Utuling

Syyskuu 2015

Tämän tutkimuksen kohde on suomen kirjoitetun yleiskielen morfosyntaktisten yhdyssanarakenteiden produktiivisuus. Tutkimuksen tärkein päämäärä on selvittää, kuinka ahkerasti erilaisia suomen kielen suomia mahdollisuuksia käytetään uusien yhdyssanojen muodostuksessa. Käytännöllistä produktiivisuutta kartoittava tutkimus täydentää kielioppien ja sanastonkuvausten antamaa kuvaa kielestä.

Tutkimuksen kohteena oleva kielimuoto on kaikille kielenkäyttäjille yhteinen kirjoitettu yleiskieli. Tutkimusaineisto koostuu 28 091 uudesta yhdyssanasta, jotka on kerätty painetun median kielestä vuosina 2000–2009. Aineiston pohjana on Kotimaisen kielten keskuksen Nykysuomen sanastotietokanta, johon poimitaan uusia ja uudella tavalla käytettyjä sanoja ensisijaisesti sanakirjatyön ja kielenhuollon tarpeisiin.

Tutkimusaihetta lähestytään useiden yhdyssanan osien muotoa, sanaluokkaa, määrää ja pituutta koskevien alakysymysten kautta. Tutkimus etenee yksittäisten muuttujien käsittelystä muuttujien keskinäisiä suhteita tarkasteleviin malleihin. Tutkimuksessa käytetään kaksivaiheista metodia: Metodien ensimmäinen askel on uudessa sanastossa havaittujen rakenteiden tyyppifrekvenssin tilastollinen analyysi. Toinen askel on varsinkin matalafrekvenssisten tai tilastollisissa analyysissä muuten poikkeavaksi osoittautuneiden rakenteiden kvalitatiivinen tarkastelu. metodi on kehitetty tätä tutkimusta varten, sillä aiemmin produktiivisuuden mittaamisessa käytetyt menetelmät eivät sellaisenaan sovi suomen kielen yhdyssanarakenteiden tutkimukseen. Tutkimusmetodien kehittäminen on tutkimuksen keskeinen tavoite.

Tutkimus osoittaa, että suomen kielen uudet yhdyssanat ovat rakenteeltaan homogeenisempiä, kuin aiempia kielenkuvauksia lukemalla voisi olettaa. Uusi suomen kielen yhdyssana on todennäköisimmin kahdesta substantiivista yhdistämällä muodostettu substantiivi, jonka alkuosa on nominatiivissa eikä kongruoi jälkiosan kanssa. Ennako-oletusta huomattavasti yleisempiä ovat myös prefiksinkaltaisella alkuosalla alkavat yhdyssanat. Genetiivialkuiset yhdyssanat puolestaan ovat ennako-oletusta harvinaisempia. Kaikki kielioppillisesti mahdolliset yhdyssanarakenteet eivät ole lainkaan produktiivisia kielenkäytön tasolla.

Tutkimus on luonteeltaan kielen rakennetta kartoittavaa perustutkimusta. Tutkimustulosten tärkeimmät sovellusalat ovat kieliteknologia ja sananmuodostuksen opetus. Tutkimus avaa useita näkökulmia jatkotutkimukselle.

Asiasanat: yhdyssanat, produktiivisuus, sananmuodostus, leksikologia, muotooppi -- kielitiede

UNIVERSITY OF TURKU

Faculty of Humanities

School of Languages and Translation Studies

Finnish and Finno-Ugric Languages

TYYSTERI, LAURA: *Aamiaiskahvilasta ötökkätarjontaan*. Suomen kirjoitetun yleiskielen morfosyntaktisten yhdyssanarakenteiden produktiivisuus

Productivity of morphosyntactic compound structures in written Finnish standard language

Doctoral Thesis, 242 p.

Doctoral Programme Utuling

September 2015

The subject of this study is the productivity of morphosyntactic compound structures in standard written Finnish. The main goal of the study is to examine the frequency of use of the range of possibilities by which new compound structures can be formed in the Finnish language. The study, which surveys practical productivity, complements the picture that grammar and lexicon descriptions give about the language.

The study focuses on a language form that is common to all Finnish language users, the standard written language. The research material consists of 28 091 new compound words, which were extracted from print media during the years 2000–2009. The material is based on the Lexical Database of Contemporary Finnish compiled by The Institute for the Languages of Finland. This neologism database is continuously updated mainly for the purposes of lexicography and language planning.

The study uses a series of sub-questions concerning the form, word class, number and length of the components of a compound. It proceeds from individual variables to models that examine mutual relations between the variables. The study employs a two-phase method. First, a statistical analysis is conducted of the type frequency of structures that have been perceived among neologisms. Second, a qualitative examination is made in particular of structures that have proved to be infrequent or otherwise atypical in the statistical analysis. The method was developed for this study, as previous methods for measuring productivity are not, as such, suitable for studying Finnish compound structures. Developing new study methods was a central objective of the study.

The study shows that new compound words in the Finnish language are more homogenous in structure than might be assumed based on earlier language descriptions. A new compound in the Finnish language is typically formed by combining two nouns, the first of which is in the nominative case and does not agree with the second noun. Compounds with a prefix-like first component also occur considerably more frequently than hypothesised. In turn, compounds in which the first component is in the genitive case occur more infrequently than hypothesised. Some of the grammatically possible compound structures proved to be entirely unproductive in practice.

The study is, by nature, basic research since it surveys the structure of the language. The most important application fields of the research results are language engineering and classroom teaching about how words are formed. The study opens up several points of view for further study.

Index terms: compound words, productiveness, word formation, lexicology, morphology
-- linguistics

Alkusanat

Kun väitöskirja on valmis, on sopiva hetki katsoa taaksepäin. Tätä tutkimusta aloittaessani kaikki tuntui selvältä. Tiesin, mitä halusin tutkia ja miten sen tekisin. Hyvin pian minulle kuitenkin valkeni, ettei esimerkiksi tutkimusaiheen rajauksesta kiinni pitäminen tai tutkimusmenetelmien valinta ollutkaan niin yksinkertaista kuin olin kuvitellut. Ilman toisten apua en olisi pystynyt tekemään kaikkia niitä suuria ja pieniä valintoja, joiden ansiosta väitöskirjani on nyt valmis.

Kiitollisena katsonkin siis nyt ympärilleni. Ilman ohjaajiani en kirjoittaisi tätä tekstiä nyt. Kunnioitettavan uran tehnyt professori Kaisa Häkkinen on ollut minulle inspiroiva esikuva aivan opintojeni alusta asti, ja ilman Kaisaa en välttämättä olisi koskaan edes kiinnostunut suomen kielen tutkimuksesta saati ryhtynyt tekemään väitöskirjaa. Ilman Kaisaa en olisi myöskään voinut saada tutkimuksen ja tieteen tekemisestä sellaista kokonais käsitystä, joka minulla on nyt.

Toisen ohjaajani Tomas Leheckan pitkä ja menestyksekkäs ura on suureksi osaksi vielä edessäpäin, ja varmaan pitkälti sen takia hän on osannut niin hyvin samais-tua väitöskirjansa kanssa tuskailevan huoliin. Tomas on myös jaksanut paneutua tutkimukseni kokonaisuuteen ja yksityiskohtiin tarkemmin, kuin kukaan saattaa ohjaajalta edes toivoa. Tomasille olen erityisen kiitollinen siitä, että hän kriittisellä hetkellä ohjasi minut keskittymään olennaiseen ja esti minua vajoamasta irrelevant-tiuden hetteiköihin tai tilastotieteellisiin soihin, joissa ilman häntä rämpisin var-masti edelleen.

Tutkimukseni esitarkastajiksi minulla on ollut ilo saada professorit Klaus Laalo ja Jarmo Jantunen. Kiitän molempia rakentavasta ja paneutuneesta palautteesta, jonka ansiosta pystyin tekemään työstäni piirun verran paremman vielä loppumet-reillä. Iloitsen myös siitä, että Klaus on lupautunut vastaväittäjäkseni.

Tutkimukseni materiaalisesta mahdollistamisesta tahdon kiittää kotiyliopis-toni lisäksi kahta organisaatiota: Koneen Säätiön myöntämien apurahojen ansiosta olen voinut keskittyä täysipäiväisesti tutkimuksen tekemiseen neljän vuoden ajan. Olen iloinen ja kiitollinen siitä, että olen saanut olla nimenomaan Koneen Säätiön suojatti, siellä arvostan paljon säätiössä tehtävää ennakkoluulotonta työtä tieteen ja taiteen hyväksi. Rahoituksen lisäksi tutkimukseni kannalta ehdottoman tärkeää on ollut Kotimaisten kielten keskukselta saamani laadukas tutkimusaineisto, jonka kaltaista en olisi pystynyt yksin kokoamaan.

Aineettoman pääoman sekä kaikenlaisen muun tuen tarjoamisesta kiitän ennen muuta oman oppiaineeni eli Turun yliopiston suomen kielen ja suomalais-ugrilaisen kielentutkimuksen väkeä. Vaikka en voi tässä mainita kaikkien nimiä, arvostan kaikkea teiltä saamaani apua, kaikkea teiltä oppimaani sekä jokaista teiltä kuulemaani kannustavaa sanaa. Erityisesti tahdon nostaa esiin monet professori Marja-Liisa Helasvuon antamat nuorelle tutkijalle korvaamattomat neuvot sekä toisilta jatko-opiskelijoita saamani vertaistuen. Tanja Toropaiselle kiitos varsinkin eräänlaisena tutorina toimimisesta jatko-opintojeni alkuvaiheessa, Ilmari Ivaskalle tilastotieteellisistä keskusteluista sekä muistiinpanojen ja koodinpätkien jakamisesta, Satu Rakkolainen-Sossalle väsymättömästä kannustuksesta ja Salla Nurmiselle yhteisistä Langnet-retkistä sekä mahdollisuudesta tuskailta yhdessä väitöskirjatyön loppuvaiheen paineita. Monista käytännön asioista huolehtimisesta tahdon kiittää Hannele Pajulaa ja väitöskirjani taittamisesta Eeva Herralaa.

Aineettoman pääoman ja uusien näkökulmien tarjoamisesta haluan kiittää myös Langnet-tutkijakoulua, johon sain kuulua suurimman osan väitöskirja-ajastani. Langnetin Kielioppi ja kielen teoria alaryhmän piirissä käydyistä avartavista keskusteluista haluan kiittää erityisesti ryhmän entisiä ja nykyisiä vetäjiä Urpoa Nikannetta, Urho Määttä ja Seppo Kittilää sekä kanssaopiskelijoitani Salla Nurmista, Milla Luodonpää-Mannia ja Elina Pallasvirtaa.

Isojen asioiden äärellä on muistettava katsoa myös lähelle. Osasta väitöskirjani syntyyn olennaisesti vaikuttaneista ihmisistä on tullut läheisiä ystäviäni. Edellä ja mainittujen lisäksi tahdon kiittää sekä tieteellisistä että muista keskusteluista ja monista hyvistä hetkistä entistä kollegaa ja nykyistä hyvää ystävääni Kaisa Lehtosaloa sekä foneetikkoystäviäni Laura Taimia ja Katri Jähiä. Kaikkia täysin tiedeyhteisön ulkopuolisia ystäviäni ja sukulaisiani tahdon – ketään erikseen nimeltä mainitsematta – kiittää siitä, että olen edelleen järjissäni ja että minulla on viimeisen kuuden vuoden aikana ollut muutakin ajateltavaa kuin työ.

Kun nyt katson vierelleni, olet onneksi siinä. Simo, kiitos kärsivällisyydestäsi, kannustuksestasi ja huolenpidostasi. Anteeksi, että olet joutunut välillisesti kantamaan niin suuren osan paineista, jotka olivat oikeastaan minun. *Nyt voimme yhdessä katsoa eteenpäin.*

Padasjoella 8.8.2015

Laura Tyysteri

Sisällys

Alkusanat.....	7
Sisällys.....	9
Taulukot.....	11
Kuviot ja asetelmat.....	12
1 Johdanto.....	13
1.1 Tutkimuksen kohde.....	14
1.2 Tutkimuksen teoreettinen tausta ja suhde aiempaan tutkimukseen.....	17
1.3 Tutkimuksen tavoitteet ja kulku.....	20
2 Yhdyssanat.....	23
2.1 Prototyyppiset yhdyssanat.....	23
2.1.1 Yhdyssanan kuvaaminen prototyypin avulla.....	24
2.1.2 Suhde johdoksiin ja perussanoihin.....	26
2.1.3 Suhde sanaliittoihin.....	29
2.2 Suomen kielen yhdyssanat.....	36
2.2.1 Yhdyssanat teoreettisena ja normatiivisena kategoriana.....	36
2.2.2 Sanatyyppin synty ja kehitys.....	39
2.2.3 Muodostaminen ja morfologinen rakenne nykysuomessa.....	41
2.3 Yhdyssanojen merkitys ja tulkinta.....	44
2.3.1 Kuuluvatko yhdyssanat leksikkoon vai kielioppiin?.....	45
2.3.2 Alkuosan muoto ja yhdyssanan merkitys.....	50
2.3.3 Kirjoitettujen yhdyssanojen prosessointi.....	52
3 Produktiivisuus.....	55
3.1 Produktiivisuus käsitteenä.....	55
3.1.1 Mikä on produktiivista?.....	56
3.1.2 Kieliopillinen ja käytännöllinen sekä diakroninen ja synkroninen produktiivisuus.....	58
3.1.3 Produktiivisuus asteilmiönä.....	60
3.2 Produktiivisuuteen vaikuttavia tekijöitä.....	61
3.2.1 Kieliopilliset, semanttiset ja leksikaaliset rajoitukset.....	61
3.2.2 Prosessointiin liittyvät seikat.....	62
3.2.3 Sosiaaliset ja tyylilliset seikat.....	65
3.3 Produktiivisuuden mittaaminen.....	67
3.3.1 Eri-ikäisten sanastojen vertailu.....	68
3.3.2 Kertamuodosteiden määrä korpuksessa (ns. Baayenin metodi).....	70
3.3.3 Kielenkäyttäjillä teetetävät tuottamis- ja päätöksentekotestit.....	72
3.3.4 Tämä tutkimus: Uudissanastossa havaittu rakenteen tyyppifrekvenssi.....	75
4 Tutkimuksen käytännön toteutus.....	78
4.1 Aineistolähde: Nykysuomen sanastotietokanta.....	79
4.2 Tutkimusaineiston rajausta.....	84
4.3 Aineiston annotointi ja hallinta.....	88
4.4 Tilastolliset menetelmät.....	93
5 Yhdyssanojen rakennepiirteiden jakauma.....	100
5.1 Pituus.....	100
5.1.1 Leksikaalisten osien määrä.....	100
5.1.2 Merkkimäärä.....	104
5.1.3 Yhteenvedo.....	108

5.2	Sanaluokka.....	109
5.2.1	Tutkimuksessa käytetty sanaluokkajako.....	109
5.2.2	Koko muodosteen sanaluokka.....	110
5.2.3	Jälkiosan sanaluokka.....	115
5.2.4	Alkuosan sanaluokka.....	118
5.2.5	Yhteenvedo.....	123
5.3	Alkuosan muoto.....	124
5.3.1	Kiteytyneen alkuosan muoto: karkea luokittelu.....	124
5.3.2	Kiteytyneen alkuosan muoto: hienojakoinen luokittelu.....	126
5.3.2.1	Nominatiivit.....	127
5.3.2.2	Epäitsenäiset leksikaaliset elementit.....	128
5.3.2.3	Yhdysosamuodot.....	131
5.3.2.4	Genetiivit.....	136
5.3.2.5	Taipumattomat ja vaillinaisesti taipuvat sanat.....	138
5.3.2.6	Muut sijamuodot.....	139
5.3.3	Yhdysosan osien kongruenssi.....	141
5.3.4	Yhteenvedo.....	147
6	Rakennepiirteet ja yhdysosan muodostaminen.....	150
6.1	Muodostustapa.....	150
6.1.1	Muodostustapa ja sanaluokka.....	152
6.1.2	Muodostustapa ja alkuosan muoto.....	155
6.1.3	Yhteenvedo.....	158
6.2	Leksikaalinen monipuolisuus.....	158
6.2.1	Leksikaalisen monipuolisuuden tarkastelussa käytetyt tunnusluvut.....	159
6.2.2	Alkuosa.....	162
6.2.3	Jälkiosa.....	171
6.2.4	Yhteenvedo.....	173
7	Rakennepiirteiden keskinäiset suhteet.....	175
7.1	Yhdysosan ja sen osien pituus.....	175
7.1.1	Leksikaalisten osien määrä ja pituus.....	176
7.1.2	Sanaluokka ja pituus.....	181
7.1.3	Yhteenvedo.....	188
7.2	Selitettävänä alkuosan muoto ja sanaluokka.....	189
7.2.1	Muiden rakennepiirteiden yhteys alkuosan muotoon.....	190
7.2.2	Eri sanaluokkiin kuuluvien sanojen yhdistyminen.....	201
7.2.3	Yhteenvedo.....	203
8	Lopuksi.....	205
8.1	Mitä tutkimuksen tulokset kertovat?.....	205
8.2	Tulosten yleistettävyyks ja arvo.....	215
8.3	Tulosten käyttö.....	218
8.3.1	Jatkotutkimus.....	218
8.3.2	Sovellusmahdollisuudet.....	221
8.4	Päätössanat.....	224
	Lähteet.....	226

Taulukot

TAULUKKO 1. Tutkimusaineiston yhdyssanojen jakauma leksikaalisten osien määrän mukaan.....	101
TAULUKKO 2. Tutkimusaineiston ja Multasen (1991) aineiston yhdyssubstantiivien jakauma leksikaalisten osien määrän mukaan.....	102
TAULUKKO 3. Tutkimusaineiston yhdyssanojen merkkimääräjakauman tilastolliset tunnusluvut.....	106
TAULUKKO 4. Tutkimusaineiston yhdyssanojen sanaluokkajakauma.....	111
TAULUKKO 5. Tutkimusaineiston yhdyssanojen alkuosan sanaluokkajakauma.....	118
TAULUKKO 6. Tutkimusaineiston yhdyssanojen alkuosan karkea morfologinen muotojakauma.....	125
TAULUKKO 7. Tutkimusaineiston yhdyssanojen nominatiivialkuosien muotojakauma.....	127
TAULUKKO 8. Tutkimusaineistossa yhdyssanan alkuosana toimivien epäitsenäisten elementtien muotojakauma.....	128
TAULUKKO 9. Tutkimusaineistossa yhdyssanan alkuosana toimivien yhdysosamuotojen muotojakauma.....	131
TAULUKKO 10. Tutkimusaineiston yhdyssanojen genetiivialkuosien muotojakauma.....	137
TAULUKKO 11. Genetiivialkuisuus koko muodosteen sanaluokan mukaan tutkimusaineistossa ja CD-PS:ssa.....	137
TAULUKKO 12. Tutkimusaineiston yhdyssanojen muussa sijassa kuin nominatiivissa ja genetiivissä olevien alkuosien muotojakauma.....	139
TAULUKKO 13. Yhdysosien kongruenssi omatunto-sanassa Google-hakujen (2/2013) mukaan.....	143
TAULUKKO 14. Osiltaan kongruoivat yhdyssanat tutkimusaineistossa sekä NS:ssa ja PS:ssa.....	146
TAULUKKO 15. Tutkimusaineiston yhdyssanojen alkuosan leksikaalinen monipuolisuus alkuosan sanaluokan mukaan.....	163
TAULUKKO 16. Tutkimusaineiston yhdyssanojen alkuosan leksikaalinen monipuolisuus alkuosan muodon mukaan.....	166
TAULUKKO 17. Tutkimusaineiston yhdyssanojen alkuosan leksikaalinen monipuolisuus koko muodosteen sanaluokan mukaan.....	170
TAULUKKO 18. Tutkimusaineiston yhdyssanojen jälkiosan leksikaalinen monipuolisuus alkuosan sanaluokan mukaan.....	171
TAULUKKO 19. Tutkimusaineiston yhdyssanojen jälkiosan leksikaalinen monipuolisuus alkuosan muodon mukaan.....	172
TAULUKKO 20. Tutkimusaineiston yhdyssanojen jälkiosan leksikaalinen monipuolisuus koko muodosteen sanaluokan mukaan.....	172
TAULUKKO 21. Yhdyssanojen luokitus merkkimäärän mukaan.....	190

Kuviot ja asetelmat

ASETELMA 1. Erialaisten aineistotyyppien soveltuvuus yhdyssanarakenteiden produktiivisuutta käsitteleviin tutkimuksiin (suomen kielestä saatavissa olevat aineistot).....	80
ASETELMA 2. Tutkimuksessa tarkasteltavat muuttujat ja niiden mahdolliset arvot.....	90
ASETELMA 3. Tutkimusaineistosta erotetut osa-aineistot.....	91
KUVIO 1. Esimerkki luokitavasta puusta.....	97
KUVIO 2. Tutkimusaineiston yhdyssanojen merkkimääräjakauma.....	105
KUVIO 3. Tutkimusaineiston yhdyssanojen ensimmäisen ja viimeisen osan merkkimääräjakauma.....	107
KUVIO 4. Yhteenvedo tutkimusaineiston alkuosan hienojakoisesta muotojakaumasta.....	148
KUVIO 5. Tutkimusaineiston yhdyssanojen muodostustavan ja sanaluokan suhde.....	152
KUVIO 6. Tutkimusaineiston yhdyssanojen muodostustavan ja alkuosan muodon suhde.....	157
KUVIO 7. Leksikaalisten osien määrän ja koko muodosteen merkkimäärän suhde tutkimusaineiston yhdyssanoissa.....	176
KUVIO 8. Leksikaalisten osien määrän ja ensimmäisen osan merkkimäärän suhde tutkimusaineiston yhdyssanoissa.....	178
KUVIO 9. Leksikaalisten osien määrän ja viimeisen osan merkkimäärän suhde tutkimusaineiston yhdyssanoissa.....	178
KUVIO 10. Eri sanaluokkiin kuuluvien yhdyssanojen merkkimäärä tutkimusaineistossa.....	181
KUVIO 11. Eri sanaluokkiin kuuluvien yhdyssanojen ensimmäisen osan merkkimäärä tutkimusaineistossa.....	183
KUVIO 12. Eri sanaluokkiin kuuluvien yhdyssanojen viimeisen osan merkkimäärä tutkimusaineistossa.....	183
KUVIO 13. Eri sanaluokkiin kuuluvien yhdyssanojen leksikaalisten osien määrä tutkimusaineistossa.....	187
KUVIO 14. Luokitava puu, jossa selitettävänä muuttujana yhdyssanan alkuosan muoto ja selittävinä muuttujina alkuosan sanaluokka sekä koko muodosteen sanaluokka ja pituus.....	192
KUVIO 15. Luokitava puu, jossa selitettävänä muuttujana yhdyssanan alkuosan muoto ja selittävinä muuttujina koko muodosteen sanaluokka ja pituus.....	196
KUVIO 16. Luokitava puu, jossa selitettävänä muuttujana yhdyssanan alkuosan muoto ja selittävinä muuttujina koko muodosteen merkkimäärä.....	199
KUVIO 17. Luokitava puu, jossa selitettävänä muuttujana yhdyssanan alkuosan sanaluokka ja selittävinä muuttujina koko muodosteen sanaluokka ja pituus.....	201
ASETELMA 4. Tiivistelmä tutkimuksen johdannossa esitettyihin alakysymyksiin saaduista vastauksista.....	206

1 Johdanto

Suomen – kuten muidenkin kielten – sanasto uusiutuu jatkuvasti. Yksi keskeinen keino uusien sanojen muodostamiseen on kahden tai useamman sanan yhdistäminen yhdyssanaksi. Suomen kielen uusista sanoista suurin osa on nimenomaan yhdyssanoja. Suomen kielessä käytettävistä yhdyssanoista puolestaan suurin osa on kahdesta substantiivista koostuvia substantiiveja, joiden alkuosa on nominatiivissa. Tätä rakennetta noudattavat esimerkiksi sellaiset sanat kuin *avaruusturisti*, *kiirekulttuuri*, *ohjelmistomato*, *rakkausloma* ja *äijäjooga*.

Rakenteeltaan tavanomaisten tapausten lisäksi suomen kielessä ovat mahdollisia myös muunlaiset yhdyssanat. Yhdyssana voi suomen kielessä edustaa mitä tahansa sanaluokkaa. Sen osina voivat toimia eri sanaluokkia edustavat sanat, pelkät sanavartalot sekä erilaiset lyhenteet ja jopa erikoismerkit. Yhdyssanan alkuosa voi olla missä tahansa sijamuodossa ja myös monenlaiset muut muodot ovat mahdollisia. Joskus alkuosa voi taipua samaan tapaan kuin yhdyssanan jälkiosa. Yhdyssanassa voi periaatteessa olla rajattomasti leksikaalisia osia. Erilaisia tavanomaisimmasta poikkeavia rakennepiirteitä sisältävät esimerkiksi sellaiset sanat kuin *@-hanke*, *arjenkarhea*, *etävastaanottaa*, *maallemuuttaja*, *kellumahiiri*, *lähialueministerivaliokunta*, *sisäkorvatyypinen* ja *ääriharvoin*.

Tämän tutkimuksen tarkoitus on selvittää, mitkä suomen kielessä mahdollisista morfosyntaktisista yhdyssanarakenteista ovat produktiivisia kielenkäytössä eli mitkä yhdyssanojen rakennepiirteet ovat kielenkäyttäjien suosiossa tällä hetkellä. Toistaiseksi ei ole olemassa tarkkaa ja tilastollisesti perusteltua tietoa siitä, kuinka ahkerasti erilaisia kielen suomia mahdollisuuksia käytetään uusien yhdyssanojen muodostuksessa. Kieliopin mahdollistamien rakenteiden kartoittamisen lisäksi sen selvittäminen, kuinka paljon rakenteita todella käytetään, on kuitenkin olennainen osa hyvää kielen kokonaiskuvausta, sillä kielioppi on kielen käytön synnyttämää ja muovaamaa. Kielen käyttö ja kielisysteemi ovat jatkuvassa vuorovaikutuksessa keskenään, eikä toista ole olemassa ilman toista (keskustelua aiheesta esim. Newmeyer 2003, 2005; Clark 2005; Laury-Ono 2005).

Kuten kahden ensimmäisen kappaleen esimerkkisanoista käy havainnollisesti ilmi, yhteen yhdyssanaan voidaan tiivistää valtavasti informaatiota. Erilaisten morfosyntaktisten yhdyssanarakenteiden suosio kertoo siitä, millä kieliopin keinoilla

tämä informaatio on kielenkäyttäjien mielestä tarkoituksenmukaisinta tiivistää. Aina uutta sanaa muodostaessaan kielenkäyttäjä valitsee tilanteeseen sopivimman rakenteen useista eri vaihtoehdoista. Kielenkäyttäjien (yleensä tiedostamatta) tekemät valinnat heijastavat esimerkiksi sitä, millaisia rakenteita pidämme riittävän helppoina sekä tuottaa että ymmärtää. Suomen kielen yhdyssanarakenteiden produktiivisuus on kieliopin ja kielen käytön vuorovaikutuksen näkökulmasta erityisen mielenkiintoinen tutkimuskohde siksi, että suomen kielioppi asettaa yhdys sanojen muodostukselle vain vähän rajoituksia, ja uutta yhdyssanaa luodessaan kielenkäyttäjä voi valita sen rakenteen periaatteessa varsin vapaasti.

Tämä tutkimus tarjoaa uutta tietoa suomen kielen sananmuodostuksesta ja yhdyssanojen rakenteesta paitsi fennisteille myös typologeille. Varsinkin muita kuin indoeurooppalaisia kieliä koskeva uusi tieto on tärkeää typologisen yhdyssanatutkimuksen edistymisen kannalta. Uuden tiedon tuottamisen lisäksi tämä tutkimus vie produktiivisuuden tutkimusta eteenpäin myös metodologisesti. Koska produktiivisuuden tutkimuksessa aiemmin käytetyt menetelmät eivät sellaisenaan sovi suomen kielen yhdyssanarakenteiden tarkasteluun, olen kehittänyt tätä tutkimusta varten uudenlaisen metodin, jonka esittely ja käytännön testaaminen on keskeinen osa tätä työtä.

1.1 Tutkimuksen kohde

Tämän tutkimuksen kohde on suomen kirjoitetun yleiskielen morfosyntaktisten yhdyssanarakenteiden produktiivisuus. Yhdyssanalla tarkoitan tässä tutkimuksessa vähintään kahden sanan muodostamaa sanaa, joka kuitenkin on yhden yksittäisen käsitteen nimitys (yhdyssanan määritelmästä ja esimerkiksi sanan ja yhdys sanan käsitteiden ongelmallisuudesta tarkemmin luvussa 2.1). Produktiivisuudella puolestaan tarkoitan kielellisen rakenteen kykyä tuottaa uusia kielenmukaisia ilmauksia (tarkemmin luvussa 3.1). Produktiivisuuden käsitteestä voidaan erottaa kaksi puolta: kieliopillinen ja käytännöllinen produktiivisuus. Kieliopillinen produktiivisuus tarkoittaa rakenteen potentiaalista kieliopillista kykyä uusien ilmausten tuottamiseen. Käytännöllinen produktiivisuus puolestaan tarkoittaa rakenteen käytännössä tuottamia uusia ilmauksia. (Kieliopillisesta ja käytännöllisestä produktiivisuudesta tarkemmin luvussa 3.1.2.) Tässä tutkimuksessa keskityn käytännölliseen produktiivisuuteen.

Käytännölliseen produktiivisuuteen päästään käsiksi tarkastelemalla rakenteen suosiota sananmuodostuksessa. Tässä tutkimuksessa keskityn senkaltaisiin yhdys sanan rakenteeseen kuuluviin morfosyntaktisiin ominaisuuksiin kuin alkuosan muoto, kokonaisuuden ja osien sanaluokka sekä osien määrä. Kussakin yhdys sanassa esiintyvät rakennepiirteet (esim. genetiivialkuisuus, adjektiivialkuisuus ja kaksisoisuus) muodostavat yhdessä yhdys sanan morfosyntaktisen kokonaisrakenteen

eli rakennetyypin¹. Tarkastelen tutkimuksessa ominaisuuksia sekä yksitellen että yhteydessä toisiinsa (luvut 5–7).

Tämä tutkimus keskittyy ensisijaisesti morfologiaan. Nimitän kuitenkin tarkastelemiani yhdyssanarakenteita morfosyntaktisiksi, sillä osalla yhdyssanojen morfologisista ominaisuuksista on myös syntaktinen puolensa. Esimerkiksi sanaluokka ei ole seurausta vain sanan omista ominaisuuksista, vaan määräytyy lopullisesti vasta kulloisessakin käyttökontekstissa. Yhdyssanan alkuosan muoto puolestaan voidaan nähdä sanan sisäisenä syntaksina². Morfologian ja syntaksin lisäksi tutkimuksen teoriaosassa esitellään myös yhdyssanojen merkityksen rakentumista ja sen suhdetta yhdyssanojen muotorakenteeseen, sillä muoto ja merkitys ovat yhdyssanoissa – kuten missä tahansa kielenilmauksissa – tiiviissä yhteydessä toisiinsa. Käytännön analyysissa semantiikka on huomioitu esimerkiksi tapauksissa, joissa morfologinen rakenne on erilainen sanan merkityksestä riippuen (vrt. vanha esimerkkisana *koululaiskuri*). Koska tämä tutkimus tarkastelee yhdyssanojen muotoa, käytän yhdyssanan yhdysosista nimityksiä *alkuosa* ja *jälkiosa* ja viittaa vain tarvittaessa niiden semanttiseen funktioon määrite- ja edusosana tai summayhdyssanan keskenään semanttisesti tasavertaisina yhdysosina (termeistä ks. ISK § 398). Käsitellessäni yhdyssanan leksikaalisia osia käytän myös nimityksiä *ensimmäinen* ja *viimeinen osa*.³

Yhdistämistä on yleensä fennistiikassa pidetty ensisijaisesti nominien muodostuskeinona, ja myös monet muita kieliä käsittelevistä tutkimuksista keskittyvät nimenomaan kaksi leksikaalista osaa sisältäviin nomineihin (toisinaan myös pelkkiin substantiiveihin). Tällöin tarkastelun kohteena on useimmiten yhdysnominien merkityksen rakentuminen. Morfosyntaktisten yhdyssanarakenteiden produktiivisuutta käsittelevässä tutkimuksessa ei kuitenkaan ole mieltä keskittyä ainoastaan tyyppillisiin rakenteisiin, sillä yksi tutkimuksen keskeisistä kysymyksistä on juuri se, kuinka ahkerasti harvinaisempia rakenteita hyödynnetään todellisessa kielenkäytössä. Tämän vuoksi käsitelen yhtä lailla esimerkiksi eri sanaluokkiin kuuluvien yhdyssanojen rakennetta.

1 Vrt. nimitys *yhdyssanatyypipi*, joka viittaa semanttisiin yhdyssanatyyppeihin. Perinteisen ja fennistisessä tutkimuksessa pääasiassa käytetyn tavan mukaan yhdyssanat on jaettu kahteen semanttiseen päätyyppiin, määritysyhdyssanoihin (esim. *kivitalo*, joka on edusosansa alakäsite) ja summayhdysanoihin (esim. *parturi-kampaamo*, jossa yhdysosat ovat semanttisesti tasa-arvoiset). Päätyypeille on lisäksi erotettu erilaisia alatyyppejä (esim. appositioyhdyssanat, bahuvihiyhdyssanat, toistoyhdyssanat, dvandvat jne.). (Ks. esim. ISK § 398–437).

2 Tässä tutkimuksessa en tarkastele yhdyssanojen sisäistä syntaktista rakennetta siinä mielessä, kuin se ymmärretään esimerkiksi useissa generativistisissa tutkimuksissa. Niissä yhdyssanan osien välille oletetaan samanlaiset syntaktissemanttiset suhteet kuin muodosteiden taustalla oleviksi ajatelluissa lauseissa (esim. Lees 1966; 1970; Saukkonen 1973; Levi 1978; Bisetto–Scalise 2005, 2009; Jussi Niemi 2009: 244–245).

3 Esimerkiksi yhdyssanan *päiväkirjamerkintä* leksikaaliset osat ovat *päivä*, *kirja* ja *merkintä*. Muodosteen ensimmäinen osa on siis *päivä* ja viimeinen *merkintä*. Muodosteen yhdysosat eli alkuosa ja jälkiosa puolestaan ovat *päiväkirja* ja *merkintä* (ISK § 403, 405, myös esimerkki ISK:n). Käytännössä tarkastelen tässä tutkimuksessa alku- ja jälkiosia vain kaksi leksikaalista osaa sisältävissä yhdyssanoissa. Enemmän kuin kaksi leksikaalista osaa sisältävien yhdyssanojen osalta käsitelen vain ensimmäisiä ja viimeisiä osia. (Ks. tarkemmin 4.3.)

Sekä produktiivisuuden tutkimusta että yhdyssanatutkimusta on toisinaan kritisoitu liiasta kirjoitettuun kieleen keskittymisestä (esim. Mäkisalo 2000a: 64; Lieber–Štekauer 2009: 7–8; Pustyl'nikov–Schneider–Wiejovski 2009: 2). Varsinkaan nykymaailmassa kirjoitettua kieltä ei kuitenkaan voida enää pitää vain puhutun kielen sekundaarina muotona ja tapana tallentaa puhetta. Kirjoitettu kieli on kielimuoto, jolla on omat erityispiirteensä ja joka nykyisenkaltaisessa kirjallisesta viestinnästä riippuvaisessa yhteiskunnassa määrittää monin tavoin sitä, miten kielenkäyttäjät kielen hahmottavat (ks. esim. Ong 2002 [1982]; Grainger–Ferrand 1994; Grainger ym. 2003). Hyvä on myös huomata, että puhutun ja kirjoitetun kielen raja ei ole selvä. Uusien medioiden ja esimerkiksi verkossa käytettävien pikaviestinten myötä raja on hämärtynyt entisestään (enemmän aiheesta esim. Crystal 2006), mutta perinteinenkin kirjoitettu kieli sisältää dialogia ja puhuttuja kommentteja.

Tässä tutkimuksessa tarkasteltava kielimuoto on suomen kirjoitettu yleiskieli. Yleisesti suomen kieltä käsittelevässä tutkimuksessa kirjoitettu yleiskieli on kaikille kielenkäyttäjille yhteisenä kielimuotona luonnollinen valinta tarkasteltavaksi vaihteeksi. Se on myös toistaiseksi ainoa kielimuoto, josta on helposti saatavilla tämän tutkimuksen kaltaisessa työssä tarvittava laaja uusien yhdyssanojen sisältävä aineisto. Tutkimuksessa käyttämäni aineisto koostuu toimitetusta painetun median kielestä (aineisto on esitelty tarkemmin luvussa 4). Aineisto on peräisin Kotimaisten kielten keskuksen (2012 asti Kotimaisten kielten tutkimuskeskus) Nykysuomen sanastotietokannasta (NSST), johon poimitaan uusia ja uudella tavalla käytettyjä sanoja ensisijaisesti sanakirjatyon ja kielenhuollon tarpeisiin. Tätä tutkimusta varten olen käynyt läpi NSST:aan 1.1.2000–1.9.2009 tehdyt poiminnat ja erottanut niistä yhdyssanat omaksi tietokannakseen. NSST:ssa ei oteta kantaa hakusanojen yhdyssanuuuteen, ja aineiston rajaus on tehty tässä tutkimuksessa esitetyn yhdyssanan määrittelyn sekä suomen kielen ortografian nojalla (tarkemmin luvuissa 2.1, 2.2.1 ja 4.2). Tutkimusaineistoon kuuluu yhteensä 28 091 yhdyssanaa. Teorialuvuissa 2 ja 3 sekä tutkimuksen käytännön toteutusta kuvaavassa luvussa 4 tutkimusaineistoon kuuluvat esimerkit on merkitty lyhenteellä TT (tutkimustietokanta). Analyysiluvuissa 5–7 ja loppuluvussa 8 kaikki esimerkit ovat tutkimusaineistosta, ellei toisin ole mainittu.

Tutkimus on luonteeltaan synkroninen ja pääasiassa kvantitatiivinen. Synkroninen näkökulma on tutkimuksessa läpikäyvä, ja sellaiset yhdyssanatutkimuksessa keskeiset käsitteet kuin (vapaa) morfeemi ja leksikaalinen elementti on tutkimuksessa määritelty synkronisesta näkökulmasta. Myös tutkimusaineiston rajaus perustuu synkroniseen näkökulmaan (tarkemmin luku 4.2). Kvantitatiivinen tarkastelu on kielen rakenteiden käytännöllistä produktiivisuutta käsittelevässä tutkimuksessa välttämätöntä, sillä hajanaiset kvalitatiiviset havainnot eivät voi kertoa produktiivisuuden asteesta (Räsänen 2010: 14). Kvantitatiivinen näkökulma mahdollistaa myös erilaisten tilastollisten menetelmien käytön (tarkemmin luku 4.4) ja sitä kautta eri rakennepiirteiden keskinäisten riippuvuussuhteiden tarkastelun. Lisäksi käytän tutkimuksessa metodina kvalitatiivista havainnointia, jota kaikki produktiivisuuden tutkimuksessa käytetyt kvantitatiiviset menetelmät vaativat tuekseen (Plag 1999: 111–115 sekä luku 3.3).

Tutkimuksessa käytetyt käsitteet on määritelty niiden esiintyessä ensimmäisen kerran. Ne peruskäsitteet, joiden määritelmää ei ole tekstissä erikseen selitetty, on tässä tutkimuksessa määritelty samalla tavalla kuin Isossa suomen kieliopissa (ISK 2004), joka on ilmestymisestään lähtien ollut Suomessa tärkeä auktoriteetti kielitieteen ja varsinkin fennistiikan käsitteiden määrittelyssä (Huumo–Ojutkangas 2005: 224).

1.2 Tutkimuksen teorettinen tausta ja suhde aiempaan tutkimukseen

Esittelen seuraavaksi lyhyesti tämän tutkimuksen aihepiiriin liittyvää aiempaa tutkimusta. Palaan tutkimusaiheen kannalta relevantteihin aiempien tutkimusten tuloksiin teorialuvuissa 2 ja 3.

Perinteisessä fennistisessä tutkimuksessa yhdyssanoja on tarkasteltu enimmäkseen kielenhuollon näkökulmasta. Erityisesti huomiota on kiinnitetty yhteen ja erikseen kirjoittamiseen (mm. Niinimäki 1991; Vanhatalo 1998 sekä lukuisat pro gradu -tutkielmat) sekä nominatiivin ja genetiivin vaihteluun yhdyssanan alkuosassa (kat-sauksia aiheeseen Vesikansa 1989: 227–237 ja T. Kolehmainen 2014: 286–299). Varsinaisten tutkimusten lisäksi näitä aiheita käsitellään myös käytännöllisesti katsoen kaikissa kielenhuolto-oppaissa ja kieliopeissa sekä lukuisissa tieteellisissä aikakauslehdissä julkaistuissa lyhyissä puheenvuoroissa (mm. Ojansuu 1908; Sadeniemi 1942; Nikanne 1971; Mielikäinen 1977; Pulkkinen 1980; Laaksonen 1984). Kielenhuoltoon keskittyvästä linjasta poikkeavat 70- ja 80-luvuilla tehdyt yhdyssanojen merkityksen rakentumista koskevat tai sitä sivuavat tutkimukset (mm. Saukonen 1973; Koski 1981; Seppänen 1981; Kosonen 1985; Räisänen 1986), Jukka Mäkisalonen (2000a, 2000b) yhdyssanoja leksikon ja kieliopin välimaastossa olevina elementteinä tarkasteleva väitöskirja, Toini Rahdun (1984) ja Terhi Malmivaaran (2004) yhdyssanojen muodostusta eri näkökulmista käsittelevät artikkelit, Kaisa Häkkisen (1987), Häkkisen ja Pia Björkqvistin (1991) sekä Tanja Vaittisen (2003, 2007) yhdyssanoja historiallisesta näkökulmasta käsittelevät artikkelit, Jussi Niemen (2009) suomen yhdyssanojen typologiaa käsittelevä artikkeli sekä Argusin, Johansen Ijäksen ja Laalon (2014) suomen-, viron- ja saamenkielisten lasten yhdyssanojen omaksumista käsittelevä artikkeli. Kielenhuollollinen näkökulma on edelleen yleisin myös yhdyssanoja käsittelevissä pro gradu -tutkielmissa, mutta viime vuosina niissä on käsitelty myös muita yhdyssanoihin liittyviä aiheita (mm. Levonen 2006; Tyysteri 2009; Lantee 2010). Keskeisenä sananmuodostuskeinona yhdistäminen ja yhdyssanat ovat saaneet huomiota myös suomen sananmuodostusta lähtökohtaisesti jostakin muusta näkökulmasta käsittelevissä tutkimuksissa (mm. P. Laine 2007; Pitkänen 2008) sekä sanastomme yleiskuvauksissa ja kieliopeissa (mm. Penttilä 1963 [1957]: § 218–224; Erkki Itonen 1966: 234–240; Vesikansa 1989). Tähän mennessä kattavimman yleisesityksen suomen kielen yhdyssanojen rakenteesta tarjoaa ISK (§ 398–437).

Eräs syy siihen, että Suomessa ei ole juuri kiinnostuttu yhdyssanojen teoreettisesta tutkimisesta, saattaa olla se, että koko yhdyssanojen kategoriaa on pidetty jonkinlaisena ortografian kautta määrittävänä itsestäänselvytenä.⁴ Kansainvälisessä yhdyssanatutkimuksessa keskeisiä tutkimuskohteita ovat olleet yhdyssanojen sisäisen rakenteen tarkastelu erityisesti semantiikan ja verbikeskeisen syntaksin näkökulmasta (esim. Lieber 2009; ten Hacken 2009; suomen osalta näitä näkökulmia ovat tarkastelleet Saukkonen 1973; Mäkisalo 2000a ja Jussi Niemi 2009). Muita yhdyssanatutkimuksessa paljon käsiteltyjä aiheita ovat yhdyssanojen kategorian määrittely (mm. Paul 1898: 301–305; Bauer 1978; Kavka 2009; Lieber–Štekauer 2009), yhdyssanojen asema syntaksin ja leksikon rajapinnalla (mm. Scalise–Guevara 2005; suomen osalta Mäkisalo 2000a) sekä pääjäsenyys (mm. Williams 1981; Zwicky 1985; Scalise 1992; Štekauer 2001). Viime vuosikymmeninä tutkimuksessa on ollut nousussa myös psykolingvistinen näkökulma esimerkiksi yhdyssanojen tunnistamiseen ja prosessointiin (katsaus aiheeseen esim. Gagné 2009; suomen osalta Hyönä–Pollatsek 1998; Pollatsek–Hyönä–Bertram 2000; Bertram–Hyönä 2003). Lisäksi huomiota on alettu kiinnittää yhdyssanojen omaksumiseen ja lapsenkielen yhdyssanoihin eri kielissä (suomen, viron ja saamen osalta Argus – Johansen Ijäs – Laalo 2014; esim. viron ja venäjän osalta Argus–Kazakovskaya 2013). Laajin tähän mennessä pelkästään yhdyssanoihin keskittyvä teos on vuonna 2009 ilmestynyt *The Oxford handbook of compounding* (OHC 2009), jossa ovat esillä sekä edellä mainitut että monet muut näkökulmat.

Ollennainen syy siihen, että kansainvälinen yhdyssanatutkimus keskittyy syntaksiin ja semantiikkaan, on tutkimuksen painottuminen englannin kieleen, jossa yhdyssanojen sisäinen morfologia ei ole relevantti tutkimuskohde. Englannin kielen ohella muita tutkimuksen kohteena olleita kieliä ovat esimerkiksi italia (varsinkin Scalise kollegoineen, esim. 1992; Bisetto–Scalise 2005, 2009; Grandi 2009)⁵, hollanti (varsinkin Booij, esim. 2007) sekä tanska (varsinkin Bauer, esim. 1978, 2009a), joissa yhdyssanojen sisäinen morfologia on rikkaampi kuin englannissa. Sekä näitä että monia muita kieliä koskevassa tutkimuksessa morfologiaan on kiinnitetty huomiota ensisijaisesti silloin, kun sen avulla on voitu osoittaa englannin kieleen

4 Myös termi *yhdyssana* on suomen kielessä niin vahvasti ortografisesti painottunut, että Mäkisaloon (2000a: 64) mukaan sen korvaamista olisi teoreettisessa tutkimuksessa syytä harkita esimerkiksi termeillä *kompositumi*, *yhdistymä* tai *sanayhdistymä*. Termi *yhdyssana* olisi hänen mukaansa varattava vain ortografiseen käyttöön. Yleensä uuden termin käyttöönottoa vakiintuneen tilalle on kuitenkin harkittava tarkoin, sillä se ei välttämättä korjaa vanhan termin leimautumisesta aiheutuneita ongelmia vaan mutkistaa tilannetta entisestään. Koska tämä tutkimus tarkastelee kirjoitettua kieltä ja koska tutkimusaineisto on rajattu osaksi ortografian kautta, käytän tässä tutkimuksessa pääasiassa perinteistä nimitystä *yhdyssana*. Tarvittaessa pyrin kuitenkin osoittamaan, milloin on kyse ortografisesta ja milloin teoreettisesti määritellystä yhdyssanasta (ks. luku 2.2.1). Sellaisissa kohdissa, joissa eronteko ei ole relevanttia, käytän nimitystä *yhdyssana*. Yksittäisistä lekseemeistä käytän tässä tutkimuksessa nimitystä *sana* silloin, kun sanoja käsitellään sanaston yksiköinä eikä sekaantumisen vaaraa esimerkiksi lekseemin ja saneen käsitteiden välillä ole.

5 Sergio Scalise johtaa kansainvälistä yhdyssanoja lähinnä typologisesta näkökulmasta tarkastelevaa MorBoComp-hanketta. Lisätietoja sivulta <http://morbocomp.sslmit.unibo.it/>. Suomen kielen edustajana projektissa on toiminut Jussi Niemi (ks. esim. Jussi Niemi 2009).

keskittyneestä tutkimuksesta johdetut universaalit virheellisiksi (mm. Bauer 1978; 2009b: 346–347; ks. myös OHC 2009: Part II). Myös yksi tämän tutkimuksen tarkoituksista on tarjota typologeille lisää tietoa eri kielten yhdyssanarakenteiden käytöstä.

Produktiivisuuden määritelmässä on perinteisesti painotettu rakenteen kieliopillista kykyä uusien ilmausten tuottamiseen. Kieliopillinen produktiivisuus oli pitkään myös useimpien alan tutkimusten lähtökohta (ks. Baayen 1992: 109). Viime vuosikymmeninä on kuitenkin alettu kiinnittää yhä enemmän huomiota myös käytännölliseen produktiivisuuteen. Produktiivisuuden kuvaaminen myös tästä näkökulmasta onkin olennainen osa hyvää kielenkuvausta, sillä kielen rakenne ja käyttö ovat jatkuvassa vuorovaikutuksessa keskenään.

Käytännöllisen produktiivisuuden tutkimuksen ovat mahdollistaneet alalla 1990-luvun taitteesta alkaen kehitellyt uudet menetöt. Aiheen suosioon on vaikuttanut tietysti myös se, että kielen käytön tutkimusta on viimeksi kuluneiden vuosikymmenten aikana ylipäänsä alettu pitää aiempaa tärkeämpänä. Aluksi alan teoriaa ja metodologiaa kehittäli erityisesti Harald Baayen kollegoineen (mm. Baayen–Lieber 1991; Baayen 1992, 1993). Myöhemmin produktiivisuuden tutkimuksen tärkeitä kehittäjiä ovat olleet muun muassa Shmuel Bolozky (1999) ja Ingo Plag (1999).

Käytännöllistä produktiivisuutta on tutkittu pääasiassa kolmella metodilla, jotka ovat 1) eri-ikäisten sanakirjojen ja muiden sana-aineistojen vertailu (tarkemmin mm. Bolozky 1999; Plag 1999; Bauer 2001: 156–158), 2) tiettyä rakennetta noudattavien kertamuodosteiden määrään perustuva alun perin Baayenin kollegoineen kehittelemä tilastollinen metodi (ns. Baayenin metodi, tarkemmin mm. Baayen–Lieber 1991; Baayen 1992, 1993) sekä 3) kielenkäyttäjillä teetetävät uuden sanaston tuottamistehtävät ja leksikaaliset päätöksentekotehtävät (*lexical decision tasks*) (tarkemmin mm. Bolozky 1999). Kertamuodosteiden määrään perustuvaa metodia on sovellettu myös lausekonstruktioiden produktiivisuuden tutkimuksessa (esim. Barðdal 2008). Monet tutkijat korostavat toisiaan täydentävien metodien soveltamista ja eri metodeilla saatujen tietojen vertailua produktiivisuuden mittaamisessa saatujen tulosten luotettavuuden takaamiseksi (Bolozky 1999: 194; Plag 1999: 177; Bauer 2001: 158–161). Esimerkiksi Bolozky (1999) käyttääkin heprean sananmuodostusta käsittelevässä tutkimuksessaan rinnan kaikkia kolmea mainittua metodia.

Suomessa produktiivisuutta on kaiken kaikkiaan tutkittu vain vähän. Useimmat tehdyt tutkimukset käsittelevät joko kieliopin mahdollistamia rakenteita (Rintala 1978; myös useimpien kielioppien ja sanastonkuvausten voidaan katsoa edustavan tätä tutkimushaaraa) tai esittävät enemmän tai vähemmän hajanaisia havaintoja eri rakenteita noudattavista kieleen syntyvistä uusista ilmauksista (mm. M. Jantunen 1997; lauserakenteiden osalta esim. Heinonen 2010: 361–364). Suomen kielen niin sanottua Baayenin metodia ovat soveltaneet suomen taivutustyyppisiä tutkineet Alexandre Nikolaev ja Jussi Niemi (2008). Eri-ikäisten sanakirjojen ja muiden sanastojen vertailua ei ole Suomessa käytetty varsinaisessa produktiivisuuden tutkimuksessa, mutta sanaston muutoksen tutkimuksessa metodia on käyttänyt

esimerkiksi Agricolan sanaston säilymistä tarkastellut Raimo Jussila (1988). Erilaisia tuottamis- ja hyväksymistestejä on produktiivisuutta koskevassa suomen kielen tutkimuksessa käytetty lähinnä taivutustyyppien tutkimuksessa (Jussi Niemi 1999; Jussi Niemi – Heikkinen – Järvikivi 2001; Nikolaev 2002; Nikolaev – Jussi Niemi 2006).

Sananmuodostuksen produktiivisuutta käsittelevissä tutkimuksissa huomautetaan usein yhdistämisen produktiivisuudesta (suomen osalta esim. Häkkinen 1990: 144; Mäkisalo 2000a: 14), mutta eri yhdyssanarakenteiden produktiivisuutta on käsitelty vain vähän. Useita tässä tutkimuksessa esiin nostetuista kysymyksistä ei ole käsitelty aiemmassa tutkimuksessa lainkaan. Esimerkiksi kvantitatiivinen näkökulma eri rakennepiirteiden yhdistymiseen uusissa yhdyssanoissa on fennistiikassa ja todennäköisesti kansainvälisestäikin aivan uusi. Kaiken kaikkiaan käytännöllisen produktiivisuuden näkökulma on noussut yhdyssanarakenteiden tutkimuksessa kansainvälisestäikin esiin vasta viime vuosina (esim. Matracki 2006; Fernández-Domínguez 2009; Ralli 2009; Stichauer 2009; Yoon 2011; Grossmann 2012). Suomessa morfosyntaktisten yhdyssanarakenteiden produktiivisuutta on toistaiseksi tarkasteltu lähinnä kieliopin suomien mahdollisuuksien kannalta, vaikka näkökulmaa ei tutkimuksissa välttämättä eksplisiittisesti mainitakaan. Esimerkiksi sananmuodostuksen kuvaukset käyttävät yleensä tätä näkökulmaa esitellessään, mikä kielessä ylipäänsä on mahdollista. Osassa tutkimuksista on esitetty myös melko hajanaisesti havaintoihin ja jonkinlaiseen mutu-tuntumaan perustuvia arvioita eri rakenteiden produktiivisuudesta. Esimerkiksi toisista yhdyssanoista johtamalla muodostettujen yhdysverbien määrän on arveltu olevan kasvussa (esim. Rahti 1984; Vesikansa 1989; Grönros 1994: 33).

Mikään produktiivisuuden tutkimuksessa aiemmin käytössä olleista metodeista ei sellaisenaan sovi suomen kielen yhdyssanarakenteiden käytännöllisen produktiivisuuden tutkimiseen (tarkemmin luku 3.3). Lisäksi produktiivisuuden tutkimuksessa on selvää tilausta myös uusille tutkimusmetodeille. Tämän takia olen kehittänyt tätä tutkimusta varten uudenlaisen tutkimusmetodin, jonka esittely ja käytännön testaaminen on yksi tämän tutkimuksen tärkeimmistä tarkoituksista. Metodi perustuu uudessa sanastossa havaittujen rakenteiden tyyppifrekvenssien tarkasteluun ja sitä seuraavaan kvalitatiiviseen havainnointiin (tarkemmin luku 3.3.4).

1.3 Tutkimuksen tavoitteet ja kulku

Tämän tutkimuksen tärkein tehtävä on selvittää, millaiset morfosyntaktiset yhdyssanarakenteet ovat nykyisessä suomen kirjoitetussa yleiskielessä produktiivisia. Käytännössä tutkimuskysymystä lähestytään useiden alakysymysten kautta. Kattavaan kielenkuvaukseen kuuluu myös kielen käytön kuvaus, ja tässä mielessä tämä tutkimus on tärkeää perustutkimusta. Tutkimuksen tuloksilla on kuitenkin myös useita soveltamismahdollisuuksia. Tärkeimmät sovellusalat ovat kieliteknologia ja sananmuodostuksen opetus. Uuden tutkimustiedon tuottamisen lisäksi

tutkimuksen tavoitteena on produktiivisuuden tutkimukseen sopivan uuden tutkimusmetodin kehittäminen.

Tutkimuksen kulku on seuraavanlainen: Tämän johdattelevan luvun jälkeen tutkimuksessa seuraa kaksi laajaa teorialukua. Niistä ensimmäinen (luku 2) käsittelee yhdyssanoja. Luvussa kiinnitetään erityistä huomiota yhdyssanan määrittelyyn, suomen kielen yhdyssanojen erityispiirteisiin sekä yhdyssanojen tulkintaan. Luvun avaa yhdyssanan teoreettisen määritelmän pohdinta (luku 2.1), sillä vain sen kautta voidaan ymmärtää useimmille suomalaisille tutumpaan oikeinkirjoituksen kautta tehtyyn määritelmään liittyviä ongelmia (luku 2.2.1). Taustoittavista luvuista toinen (luku 3) käsittelee tutkimuksen toista keskeistä käsitettä, produktiivisuutta. Myös tämä luku lähtee liikkeelle tarkasteltavan käsitteen määrittelystä. Lisäksi luvussa keskitytään erityisesti yhdyssanarakenteiden produktiivisuuteen vaikuttaviin tekijöihin sekä produktiivisuuden mittaamisessa käytettyihin metodeihin. Tässä tutkimuksessa käytetty metodi esitellään luvussa 3.3.4.

Ennen analyysia esittelen tutkimuksen käytännön toteutuksen: tutkimusaineiston, sen käsittelyn ja käytetyt tilastolliset menetelmät (luku 4). Varsinainen analyysi käynnistyy erilaisten yhdyssanan rakennepiirteiden tarkastelulla (luku 5). Luvussa esitellään kaikki tutkimuksessa tarkasteltavat muuttujat sekä niiden jakauma aineistossa. Tutkimuksessa tarkasteltavat muuttujat liittyvät yhdyssanan pituuteen, sen osien sanaluokkaan sekä alkuosan muotoon. Luvun 5 pääkysymykset ovat:

Pituus

- 1) Kuinka paljon leksikaalisia osia suomen kielen uudet yhdyssanat sisältävät? (luku 5.1.1)
- 2) Kuinka paljon merkkejä suomen kielen uudet yhdyssanat sisältävät? (luku 5.1.2)

Osien sanaluokka

- 3) Mitä sanaluokkaa suomen kielen uudet yhdyssanat ja niiden jälkiosat edustavat? (luvut 5.2.2 ja 5.2.3)
- 4) Mitä sanaluokkaa suomen kielen uusien yhdyssanojen alkuosat edustavat? (luku 5.2.4)

Alkuosan muoto

- 5) Missä muodossa suomen kielen uusien yhdyssanojen alkuosat ovat? (luvut 5.3.1 ja 5.3.2)
- 6) Taipuvatko suomen kielen uusien yhdyssanojen alkuosat? (luku 5.3.3)

Luvun 5 päätarkoitus on antaa yleiskuva tutkimusaineistosta ja löytää kvantitatiivisen analyysin avulla erityisen yleiset ja harvinaiset rakennepiirteet. Harvinaiset rakenteet eivät välttämättä ole rakenteina produktiivisia, vaan niiden esiintyminen uudessa sanastossa voi olla seurausta esimerkiksi yksittäisiin sanoihin perustuvista analogioista (esim. *alitamunta* > *ylitamunta* [TT]). Yhdyssanojen muodostuksessa produktiivinen voi rakenteen sijaan olla myös jokin tietty yhdyssanojen osana usein toistuva lekseemi (esim. *-in*-päätteinen monikon genetiivi on yhdyssanojen

alkuosana produktiivinen vain tiettyjen sanojen yhteydessä, esim. *kansain-*, *vanhempain-*⁶). (Ks. tarkemmin luku 3.3.1.) Varsinkin harvinaiset rakenteet vaativat siis kvantitatiivisen tarkastelun lisäksi myös huolellista kvalitatiivista analyysia.

Luvussa 6 tarkastelen yhdyssanan rakennepiirteiden yhteyttä sanojen muodostustapaan ja leksikaaliseen monipuolisuuteen. Myös tässä luvussa menetelminä ovat sekä kvantitatiivinen että kvalitatiivinen analyysi. Luvun 6 pääkysymykset ovat:

- 1) Millä tavalla suomen kielen uudet yhdyssanat muodostetaan? (luku 6.1)
- 2) Kuinka monipuolisesti suomen kielen uusissa yhdyssanoissa käytetään eri leksemejä? (luku 6.2)

Luku 6 jatkaa luvussa 5 aloitettua sellaisten rakenteiden etsintää, joita noudattavia muodosteita syntyy lisää mutta jotka eivät kuitenkaan ole (yhdyssana)rakenteina produktiivisia. Esimerkiksi ennemmin johtamalla kuin yhdistämällä uusia muodosteita keräävät rakenteet on katsottava produktiivisiksi ennemmin johtamisen kuin yhdyssanojen muodostuksen näkökulmasta. Yhdyssanojen leksikaalisen monipuolisuuden tarkastelu puolestaan auttaa havaitsemaan rakenteet, joiden suosio on ennemmin seurausta joidenkin yksittäisten lekseemien kuin rakenteen itsensä suosiosta.

Luvussa 7 syvennetään tietoa yhdyssanojen rakennepiirteiden käytöstä sananmuodostuksessa tarkastelemalla piirteiden keskinäisiä vaikutussuhteita sekä tilastollisen kuvailun että testauksen avulla. Luvun tärkein tavoite on selvittää, mitkä yhdyssanojen rakennepiirteet esiintyvät toisia todennäköisemmin yhdessä. Luvun moninaiset tutkimuskysymykset voidaan tiivistää yhteen kysymykseen: Millaisia keskinäisiä vaikutussuhteita suomen kielen uusien yhdyssanojen rakennepiirteiden välillä on? Konkreettisemmin luvussa vastataan esimerkiksi senkaltaisiin kysymyksiin kuin:

- 1) Ovatko suomen kielen uusien yhdyssanojen leksikaaliset osat yhtä pitkiä riippumatta siitä, montako niitä sanassa on? Ovatko eri sanaluokkiin kuuluvat yhdyssanat eripituisia? (luku 7.1)
- 2) Millaisissa muodoissa eri sanaluokkiin kuuluvat yhdyssanan alkuosat esiintyvät uudessa yhdyssanastossa? Millainen on eri sanaluokkiin kuuluvien yhdyssanojen alkuosan sanaluokka- ja muotojakauma? Ovatko esimerkiksi yhdysadjektiivit yhdysubstantiiveja useammin genetiivialkuisia? (luku 7.2)

Tutkimuksen päättää yhteenveto tuloksista (luku 8). Päätulosten yhteen kokoamisen lisäksi hahmottelen tulosten sovellusmahdollisuuksia sekä jatkotutkimuksen suuntaa. Pohdin luvussa 8 myös tutkimuksen onnistumista sekä tulosten arvoa ja yleistettävyyttä.

⁶ Muodon voidaan myös ajatella kehittyneen tai olevan kehitymässä eräänlaiseksi vain tiettyjen sanojen yhteydessä esiintyväksi yhdysosamuodoksi, sillä se on nykykielessä vanhahtava ja kaiken kaikkiaan hyvin harvinainen.

2 Yhdyssanat

Tämä luku on ensimmäinen tutkimustani taustoittavista teorialuvuista. Luku koostuu kolmesta alaluvusta, joista ensimmäisessä (2.1) määrittelen yhdyssanan käsitteen. Koska monien muiden kielellisten luokkien tapaan myös yhdyssanojen kategorian rajat ovat häilyvät, käytän määrittelynä apuna prototyypin käsitettä. Toisessa alaluvussa (2.2) käsittelen suomen kielen yhdyssanoja fennisteille yleisesti tutuimpien oikeinkirjoituksen, kielihistorian ja sananmuodostuksen näkökulmasta. Kolmas alaluku (2.3) käsittelee yhdyssanojen merkityksen rakentumista ja tulkin-taa.

2.1 Prototyyppiset yhdyssanat

Tieteellisessä luokittelussa käytettiin pitkään niin sanottua klassisen kategorian periaatetta, jossa luokat ovat selvärajaisia ja täsmällisesti määriteltyjä. Selvärajaisilla kategorioilla operoiminen on yksinkertaista ja pitää asiat yksiselitteisinä, mutta todellisessa maailmassa ja ihmisen kognitiossa, joita tieteen pitäisi pystyä kuvaamaan, luokkien rajat ovat kuitenkin sumeita (esim. P. Leino 2002a [1987]: 527–530; Onikki-Rantajääskö 2001: 33; klassisten kategorioiden kritiikkiä mm. Rosch 1977; Lakoff 1987). Tässä tutkimuksessa kuvaan yhdyssanojen luokkaa prototyypin käsitteen avulla. Prototyyppisen yhdyssanan kuvauksen pohjana käytän aiemmassa tutkimuksessa yhdyssanojen täsmällisen kategorian kriteereinä esitettyjä piirteitä. Esimerkkeinä käytän sekä tyypillisiä että epätyypillisiä tapauksia. Prototyypin kuvauksen tarkoitus on tässä tutkimuksessa paitsi esitellä tutkimuskohde myös täsmentää fennistiikassa käytettyä yhdyssanan määritelmää. Lisäksi kuvaus toimii aineistonrajauksen pohjana varsinkin suhteessa johdoksiin (aineistonrajauksesta tarkemmin luku 4.2).

Tämän luvun ensimmäisessä alaluvussa (2.1.1) pohdin prototyypin yhdyssanan kuvauksessa käytettävien piirteiden valintaa. Toisessa ja kolmannessa alaluvussa kuvaan prototyyppistä yhdyssanaa vertaamalla yhdyssanoja johdoksiin (2.1.2) ja sanaliittoihin (2.1.3). Sekä toisessa että kolmannessa alaluvussa päähuomio on

suomen kielen yhdyssanoissa, joita vertailen sekä kansainvälisen että fennistisen yhdyssanatutkimuksen pohjalta hahmoteltuun prototyyppiin.

Yhdyssanojen ja sanaliittojen erottaminen on saanut Suomessa aiemmin paljon enemmän huomiota kuin yhdyssanojen ja johdosten erottaminen, sillä fennistinen yhdyssanatutkimus on pitkään ollut kielenhuollollisesti orientoitunutta ja yhdyssanojen ja sanaliittojen erottamista on pidetty tärkeänä nimenomaan oikeinkirjoituksen vuoksi. Yhdyssanan ja sanaliiton erottaminen on ollut myös kansainvälisessä yhdyssanatutkimuksessa suuremman mielenkiinnon kohteena, sillä näiden käsitteiden eroja pohdittaessa päästään käsiksi syntaktisten ja morfologisten kielenilmiöiden rajankäyntiin.

2.1.1 Yhdyssanan kuvaaminen prototyypin avulla

Yhdyssana on perinteisesti määritelty sanaksi, joka koostuu kahdesta tai useammasta sanasta, mutta on silti yhden yksittäisen käsitteen nimitys (mm. Paul 1898: 301–305; Bauer 1978: 54; ISK § 398; Lieber–Štekauer 2009: 4–8). Tällainen määritelmä on riittävä esimerkiksi perusopetuksen tarpeisiin. Tarkkaan rajattujen klassisten kategorioiden näkökulmasta siinä on kuitenkin perustavanlaatuisia ongelmia, joista suurimmat ovat seurausta epämääräisestä sanan käsitteestä⁷. Ensinnäkin useissa kielissä – suomi mukaan lukien – yhdyssanoja voidaan muodostaa myös elementeistä, jotka eivät ole itsenäisiä sanoja vaan ennemmin jonkinlaisia juuri- tai tyvielementtejä (Lieber–Štekauer 2009: 4–6; ks. myös Sadeniemi 1949: 43–48; Penttilä 1963 [1957]: § 129, 221). Toiseksi määritelmän avulla ei voida erottaa toisistaan yhdyssanoja ja sanaliittoja, sillä yhdyssanojen tapaan monet sanaliitot⁸ ovat yhden yksittäisen käsitteen leksikaalistuneita nimityksiä (esim. Nenonen 2002: 25–32; Kavka 2009).

Sanan käsitteen epämääräisyyden lisäksi yhdyssanan tarkan määrittelyn tekee vaikeaksi se, että monet yhdyssanojen ominaisuuksista ovat ainakin jossain määrin kieli- tai kieliryhmäkohtaisia (Lieber–Štekauer 2009: 4–14). Koska suurin osa tehdystä teoreettisesta yhdyssanatutkimuksesta käsittelee pääasiassa englantia ja muita indoeurooppalaisia kieliä, aiemmassa tutkimuksessa yhdyssanojen kategorialle esitetyt kriteerit soveltuvat ensisijaisesti vain näihin kieliin.

Koska yhdyssanan käsite tuntuu pakenevan tarkkoja määritelmiä, jotkut ovat kiistäneet koko kategorian olemassaolon (esim. Marchand 1967; Nieminen–Lantee 2011). Tarkkarajaisen ja täsmällisen määritelmän puuttuminen ei kuitenkaan tar-

⁷ Sanaa ei ole pystytty määrittelemään tyydyttävästi siitä huolimatta, että se on eräs kielitieteen tärkeimpiä käsitteitä (esim. Lieber–Štekauer 2009: 4). Niin sanottujen primitiivikäsitteiden on muillakin aloilla todettu olevan vaikeita määriteltäviä (Karlsson 1994: 75).

⁸ Oikeastaan täsmällisempi nimitys leksikaalista yksikköä muistuttaville kahden sanan yhtymille olisi *liittosana* (Erkki Itkonen 166: 236; Häkkinen 1990: 147), sillä termillä *sanaliitto* voidaan viitata mihin tahansa kahden tai useamman peräkkäin esiintyvän sanan liittoon (Häkkinen mp.). Tässä tutkimuksessa en kuitenkaan tee eroa liittosanojen ja muiden sanaliittojen välille.

koita, että itse ilmiötä ei olisi olemassa. Eräs tapa⁹ kuvata luokkien sumearajaisuutta on määritellä luokat prototyypin kautta. Prototyypiajattelussa luokkakäsenyys nähdään ehdottoman sijaan asteittaisena, ja prototyypin kautta määriteltyyn luokkaan kuuluu sekä prototyyppeisiä että perifeerisiä jäseniä. Prototyyppeisiä ovat luokan luonteenomaisimmat esimerkit tai luokan kaikkein selvimmät jäsenet, eräänlaiset mallijäsenet. (Langacker 1991: 17; P. Leino 2002a [1987]: 534; Onikki-Rantajääskö 2001: 33). Yhdessä prototyyppeiset ja perifeeriset jäsenet muodostavat verkoston, jossa prototyyppi on tärkeä solmukohta tai keskusjäsen.¹⁰ Verkoston muut jäsenet liittyvät prototyyppeihin erilaisten suhdejärjestelmien kautta. Luokkien jäsenet eivät muodosta tiukasti määriteltyjä joukkoja, koska ei ole olemassa sellaista prototyypistä poikkeamisen astetta, jonka ylityttyä yhtäläisyyttä prototyypin kanssa olisi ehdottoman mahdotonta havaita. Voidaan sanoa ainoastaan, että samankaltaisuus prototyypin kanssa nostaa todennäköisyyttä, että entiteetti luokitellaan prototyypin kanssa samaan luokkaan kuuluvaksi. (Langacker 1987: 380, 1999: 17.)

Kielitieteessä prototyypin käsitteen keskeisin käyttöalue on kognitiivinen semantiikka, mutta se soveltuu myös muiden kielen osa-alueiden kuvaamiseen. Sen avulla onkin alettu kuvata myös abstrakteja kielellisiä luokkia kuten sanaluokkia tai lauserakenteita (Taylor 1998; Rissanen 1998; Croft 2003: 185; Alho-Kauppinen 2008: 111–127). Prototyypin avulla voidaan kuvata myös yhdyssanojen kategoriaa. Käytännössä esimerkiksi aineiston rajauksessa joudutaan kuitenkin tekemään myös tarkkoja rajanvetoja (tämän tutkimuksen osalta luku 4.2).

Prototyyppeisen yhdyssanan kuvauksessa käytettävien piirteiden valinta ei ole yksinkertaista. Koska osa piirteistä on kielikohtaisia ja lähes kaikkien piirteiden prototyyppeisyysaste vaihtelee kielittäin, eri kielten tutkijat ovat korostaneet osin eri piirteiden tärkeyttä (Lieber-Štekauer 2009: 4–14). Osa prototyypin kautta määrittävää luokkaa kuvaavista piirteistä voi olla toisia keskeisempiä, ja olio, jolla on jokin keskeinen luokkaa määrittävä ominaisuus, voidaan katsoa luokkaan kuuluvaksi vaikka siltä puuttuisi monia muita luokkaan yhdistettyjä piirteitä. Esimerkiksi Matthews (1991: 93–101) pitää tärkeimpinä yhdyssanan ominaisuuksina sellaisia morfologisia piirteitä, joiden puuttuminen osoittaa rakenteen selvästi periferiseksi. Myös fennisisessä tutkimuksessa on huomautettu, että esimerkiksi sellaiset sanaparit, jonka osista ensimmäinen on nominatiivissa oleva substantiivi tai yhdyssamuoto, ovat luontaisesti yhdyssanoja niiden muista ominaisuuksista riippumatta,

9 Toinen, joskin harvemmin kielitieteessä käytetty tapa on ns. sumeiden joukkojen teoria (esim. Koskinen 1992). Vanhastaan eräs tapa välttää sumearajaisuuden ongelmaa on ollut jättää tarkkarajaisiin kategorioihin sopimattomat tapaukset tutkimuksen ulkopuolelle (ks. esim. Mikone 2002: 37–43). Jos kielitieteen kuitenkin tahdotaan huomioivan ihannetapausten lisäksi myös hankalat rajatapaukset, luokkien sumearajaisuus on hyväksyttävä ja pystyttävä kuvaamaan jollain tavalla.

10 Luokan jäsenet voivat liittyä toisiinsa myös ketjuna (Onikki 2000: 93). Tällainen käsitys tuo prototyyppeisyyden ajatuksen lähelle Wittgensteinin (1981 [1953]: 65) ajatusta perheyhtäläisyydestä. Perheyhtäläisyysajattelun mukaan ei ole olemassa mitään kaikkia luokan jäseniä yhdistävää ominaisuutta vaan erilaisia päällekkäisten ja ristikkäisten yhtäläisyyksien verkostoja. Vaikka perheyhtäläisyysteoria usein mainitaan prototyyppeisyyden yhteydessä, teorioissa ei kuitenkaan ole kyse samasta asiasta: perheyhtäläisyysteoriasta puuttuu prototyypin käsite, joka taas on prototyypiajattelun perusta.

sillä niitä vastaava sanaliitto ei ole suomen kielessä mahdollinen syntaktinen rakenne (ks. esim. ISK § 400). Näin ollen esimerkiksi ilmaukset *sammakkolapsi* ja *perhoslapsi* (esimerkit keksittyjä) ovat yhdyssanoja, vaikka ne eivät ole minkään vakiintuneen käsitteen nimityksiä.

Mäkisalon (2000a: 102) mukaan myös genetiivialkuisuutta voidaan pitää osoituksena muodosteen yhdyssanuudesta sellaisissa tapauksissa, joissa genetiivialkuiselle ilmaukselle voidaan löytää nominatiivilla tai yhdyssanamuodolla alkava minimipari (*rotanmyrkky – hyönteismyrkky*). Näkemys perustuu kuitenkin analyysiin, joka olettaa yhdyssanojen taustalle rajallisen määrän semanttisia syvärakenteita eikä ota huomioon esimerkiksi yksittäisten ilmausten vakiintuneisuuden astetta (esim. *ihmislapsi – eläinlapsi – humanoidin lapsi*). Lisäksi Mäkisalon (mp.) mainitseman esimerkin kaltaisia minimipareja on hyvin harvassa.

Morfologiset piirteet kuten alkuosan muoto voivat suomen kielessä osoittaa sanan väistämättä yhdyssanaksi. Kuitenkaan kaikki perinteisesti yhdyssanoina pidetyt muodosteet eivät täytä yhdyssanalle asetettuja morfologisia kriteerejä. Yhdyssanojen luokkaa ei näin ollen voi rajata vain morfologisten kriteerien mukaan, eivätkä pelkät morfologiset kriteerit riitä myöskään prototyypin yhdyssanan kuvaukseen.

Matthewsin (1991: 93–101) mukaan semanttisten piirteiden kautta rajattu yhdyssanojen luokka on ainakin englannissa muiden kriteerien rajaamaan ryhmään verrattuna liian suuri ja fonologisten kriteerien mukaan rajattu puolestaan liian pieni. Suomen kielestä ei toistaiseksi ole tutkittua tietoa siitä, millaisena yhdyssanojen luokka näyttäytyisi, jos se rajattaisiin vain semanttisten tai fonologisten kriteerien mukaan. Selvää kuitenkin on, että semanttisten ja fonologisten ominaisuuksien tarkastelu on paikallaan myös suomen kielessä varsinkin silloin, jos rajaavat morfologiset kriteerit eivät täyty.

Sekä morfologisesti, syntaktisesti, semanttisesti että tavanomaisesti lausuttuna myös fonologisesti prototyypin yhdyssanoja ovat esimerkiksi *kivitalo* ja *kielitiede*. Prototyypin yhdyssanan piirteitä ovat seuraavat: 1) muodoste koostuu itsenäisistä lekseemistä, 2) muodosteen osien välillä kulkeva raja vastaa syntaktista rajaa, 3) muodosteella on puheessa vain yksi pääpaino (ensimmäisen osan ensimmäisellä tavulla) 4) muodostetta vastaava sanaliitto ei ole mahdollinen syntaktinen rakenne, 5) muodoste on eriytynyt osistaan semanttisesti ja kiteytynyt yksittäisen käsitteen nimitykseksi ja 6) muodoste on morfologisesti eheä ja koskematon kokonaisuus. Käyn prototyypin yhdyssanan piirteitä tarkemmin läpi kahdessa seuraavassa luvussa, joissa vertaan yhdyssanoja johdoksiin (2.1.2) ja sanaliittoihin (2.1.3).

2.1.2 Suhde johdoksiin ja perussanoihin

Prototyypiset yhdyssanat koostuvat kahdesta itsenäisestä lekseemistä (esim. *talvi-rengas*). Johtamisen ja yhdistämisen tärkeimpänä erona onkin perinteisesti pidetty sitä, että yhdyssanassa yhdistyy kaksi tai useampia vapaita morfeemeja kun taas johdoksessa vapaita morfeemeja on vain yksi (esim. *talvinen*, poikkeuksena yhdyssanasta johdetut uudet yhdyssanat, esim. *talvirengasmainen*; vapaan morfeemin

määrittelystä ks. esim. Karlsson 1994: 91–92). Kuitenkin yhdyssanojen kategoriaan luetaan myös sellaisia sanoja, joiden osat eivät voi esiintyä itsenäisinä leksemeinä. Tällaisten osien itsenäisyyden kannalta periferisten yhdyssanojen osat ovat aina sidottuja johonkin toiseen elementtiin (esim. *kaukojuna*; toisaalta epäitsenäisiä elementtejä sisältävä yhdyssana on sikäli prototyypin, että sitä vastaava sanaliitto ei ole mahdollinen, ks. tarkemmin seuraava luku 2.1.3). Joskus jopa yhdyssanan molemmat osat voivat olla epäitsenäisiä elementtejä (esim. *alakerta*¹¹). Esimerkiksi Penttilän (1963 [1957]: § 219) mukaan yhdyssana koostuukin vapaiden morfeemien sijaan kahdesta tai useammasta sanasta, joiden ”aikaisempi itsenäinen käyttö on todettavissa tai ainakin suhteellisen helposti selvitettävissä” (Sadaniemi [1949: 43] ja Penttilä [mt. § 129, 221] käyttävät yhdyssanan osista nimitystä *tyvi*). Erkki Itkonen (1966: 234) puolestaan toteaa yhdyssanan koostuvan elementeistä, jotka ”palautuvat itsenäiseen sanaan”¹².

Yhdyssanan epäitsenäisen leksikaalisen osan erottaminen johtimesta ei ole aina yksiselitteistä. Selkeä periaatteellinen ero yhdyssanojen ja johdosten välillä on niiden osien välillä kulkevan rajan luonne. Prototyypin yhdyssanojen osien välinen raja vastaa syntaktista sanarajaa, sillä esimerkiksi vokaalisoinnutus ei ulotu osien rajan yli (vrt. *päiväkoti* ja *tanssijat*; *näyttelijätär*). Yhdyssanojen lisäksi vokaalisointu rikkoutuu kuitenkin myös monissa lainasanoissa (esim. *analyysi*, *primääri*, *syntaksi* jne.). Toisaalta taas jotkin itsenäisistä leksemeistä kohti johdinta kehittyneet elementit käyttäytyvät muuten johtimen tavoin mutta eivät noudata vokaalisointua (esim. *tämän lajinen* > *tällainen*, tosin puheessa myös *tälläinen*). (Ks. myös Koski 1982: 11, 66–68.)

Yhdyssanan osien välinen raja vastaa syntaktista rajaa myös sikäli, että prototyypisissä yhdyssanoissa itsenäiset sanat yhdistyvät joko nominatiivisijaisina tai taivutuspäätteineen, kun taas johdoksissa johtimet liittyvät suoraan sanavartaloon (tosin nominatiivi voi olla identtinen sanan vartalon kanssa, ks. myös Rintala 1972: 25). Yhdyssanoissa säännönmukainen poikkeus tästä ovat yhdyssanan alkuosana toimivat *-nen*-loppuisten sanojen yhdysosamuodot, jotka ovat sanan konsonantivartalon kaltaisia (esim. *kärpäslätkä*). Myös muiden kuin *-nen*-loppuisten sanojen vartalot ovat melko yleisiä yhdyssanojen alkuosina (esim. *suurkaupunki*, *sikolätti*¹³). Muiden sidonnaisten morfeemien tapaan myös yhdysosamuodot toisaalta lähentävät rakennetta johdoksiin mutta toisaalta osoittavat sen sikäli selvästi yhdyssanaksi, että vastaava sanaliitto ei ole mahdollinen (yhdyssanoja ja sanaliittoja erottavista piirteistä tarkemmin seuraavassa luvussa 2.1.3).

11 Ilmauksen merkitys on ’alempi kerros’, eivätkä osat *ala-* ja *-kerta* eivät esiinny siinä samassa merkityksessä kuin itsenäisinä sanoina.

12 Esimerkiksi Niinimäki (1991: 24–25) laskee yhdyssanoiksi kaikki sellaiset sanat, joiden osat ovat jollain tavalla identifioitavissa itsenäisiin sanoihin (esim. *televisio*). Ratkaisu sopii Niinimäen käytännönläheiseen tutkimukseen, joka käsittelee ensisijaisesti oikeinkirjoitusta. Teoreettisesta näkökulmasta mahdollisesti kaksiosaisiksi hahmottuvat mutta kieleen kokonaisuuksina lainatut ilmaukset (esim. *papukaija*) eivät ole yhdyssanoja.

13 *A*-loppuisista substantiiveista vanhastaan käytetty *-o-* tai *-i-*loppuinen muoto vastaa käytännössä monikkovartaloa (esim. Hämäläinen 1883: 122–124).

Hyvä esimerkki johtimen ja itsenäisen sanan rajalla häilyvästä suomen kielen elementistä on Päivi Rintalan (1972) tutkima moderatiivinen *-läntä*-aines¹⁴ (esim. *vähäläntä*), joka on alun perin ollut todennäköisesti *-la*- ja *-ntA*-johtimien yhtymä. Viimeistään 1500- ja 1600-luvulla *-läntä*-aineksella on ollut myös yhdyssosan piirteitä ja ainesta on käytetty myös genetiivialkuisten yhdyssanojen edusosana (esim. *laihanläntä*). Rinnalla on kuitenkin esiintynyt myös johtimenkaltaista käyttöä (esim. *laihalanta*, *vähäläntä*). Nykykielessä *läntä* muistuttaa johdinta sikäli, että se on epäitsenäinen morfeemi, joka esiintyy vain toisten morfeemien yhteydessä. Myös sen yleisluonteinen moderatiivinen merkitys saa sen muistuttamaan johdinta. Johtimista sitä puolestaan erottaa se, että se liittyy nykykielessä nominatiiviin tai genetiiviin eikä sanavartaloon kuten johtimet yleensä. Se ei myöskään enää mukaudu vokaalisointuun vaan on nykysuomessa aina etuvokaalinen. Useat varhaiset sanankirjojen ja kielioppien tekijät ovat esittäneet *-läntä*-aineksen joko itsenäisenä sanana tai epäitsenäisenä yhdyssosana. Vasta 1900-luvun kirjallisuudessa on noussut vallalle käsitys, jonka mukaan aines olisi ensisijaisesti johdin, jolla kuitenkin on myös itsenäistä leksemiä muistuttavia ominaisuuksia. (Rintala 1972: 13–25, 244–246.) Rintalan (mts.) mukaan *-läntä*-aines voi toimia sekä johtimena että yhdyssosana, ja omassa tutkimusaineistossaan hän luokittelee aineksen sen kulloisenkin käytön mukaan.

Kuten *-läntä*-aines osoittaa, kielen systeemi ei sinänsä edellytä elementtien kuumista tiukkarajaisiin kategorioihin vaan samaa elementtiä voidaan käyttää useissa eri kategorioissa. Prototyypppeihin perustuva kielenkuvaus ei edellytä edes yksittäisten käyttötapojen luokittelua yhteen tiettyyn kategoriaan kuuluvaksi. Tutkimusaineiston rajauksessa tarvitaan kuitenkin rajanvetoja, ja vaikka itsenäisen sanan ja johtimen välimaastoon jäävät elementit eivät ole kielenkäytössä yleisiä, niiden luokittelussa tehdyt ratkaisut ovat huomionarvoisia tutkimusaineiston koostamisen kannalta. Rintalan (1972: 25) tapaan olen omassa tutkimuksessani tehnyt rajanvedot tapauskohtaisesti (aineistonrajauksesta tarkemmin luvussa 4.2).

Omassa tutkimusaineistossani johtimen ja yhdyssosan välimaastossa olevista elementeistä ongelmallisimpia luokiteltavia ovat niin sanotut vierassananlopudet, jotka käyttäytyvät monella tapaa sekä johtimen että itsenäisen morfeemin tavoin (ks. myös ISK § 152; kansainvälisessä tutkimuksessa tällaisia aineksia on nimetty sanoilla *semiword* ja *affixoid*, Schmidt 1987; Kenesei 2007). Esimerkiksi *-nomi*- ja *-nautti*-elementtejä voidaan käyttää suomen sanamuodostuksessa sekä johtimen että epäitsenäisen leksikaalisen elementin tavoin (esim. suomalaiselle avaruuslentäjälle ehdotetut nimitykset *avarnautti* [esim. FSNordic.net: Tyhmiä kysymyksiä], *tähtinautti* [Suomi24: Kosmonautti]). Täysin itsenäisinä sanoina ne esiintyvät harvoin, ja silloinkin käyttö on usein ellipsinomaista (esim. 1). (Tyysteri 2011b.) Samaan tapaan ellipsinomaisesti voidaan toisinaan käyttää myös kotoisia johtimia (esim. 2).

14 Havainnollisuuden vuoksi käytän nykykielenmukaista merkintätapaa *-läntä* myös 1500- ja 1600-lukuja kuvaavan *-lAntA*-merkintätavan sijaan.

- (1) jos joku suomalainen menee avaruuteen niin mikäs *nautti* se sitten on?
[Napsu: Mitä eroa on kosmonautilla ja astronautilla?]
- (2) – – siellä olivat läsnä laulajattaret, näyttelijättäret ja muut *taret*¹⁵ [Penttilä 1963 [1957]: § 217]

Yhdyssanan ja johdoksen tai jopa yhdyssanan ja perussanan välinen raja voi hämärtyä myös silloin, jos alun perin yhdistämällä muodostetun sanan rakenne muuttuu ajan kuluessa (Erkki Itkonen 1966: 239). Tällaiset yhdyssanat ovat muuttuneet yksimorfeemisen sanan tai johdoksen kaltaisiksi, eikä niitä voi enää tunnistaa yhdyssanoiksi ilman kielihistoriallista tietoa. Hämärtyneet muodosteet voivat jopa mukautua vokaalisointuun, kuten on tapahtunut muodosteen *iankaikkisuus* alkuosalle, joka on alkujaan *ikä*-sanan genetiivimuoto (NES s. v. *iankaikkisuus*). Alun perin yhdistämällä syntyneet sanat voivat kulua äänteellisesti myös muilla tavoilla niin, että kielenkäyttäjät ei enää tunnista niiden yhdyssanuutta. Esimerkiksi partikkelikieli *sentään* on lyhentynyt muoto sanayhtymästä *sen tähden*. Virossa ja joissakin suomen murteissa vuotta merkitsevä *aasta* on puolestaan tiivistynyt yhdyssanasta *ajastaika*, joka puolestaan lienee saanut alkunsa sanaliitossa *ajasta aikaan* (SMS s. v. *ajastaika*). (Erkki Itkonen 1966: 239–240.) Tarkkoja rajanvetoja vaativissa tilanteissa (esim. aineistonrajaus) lopullinen päätös siitä, katsotaanko äänteellisesti kulunut ilmaus yhdyssanaksi vai ei, on tehtävä sen mukaan, onko tarkastelunäkökulma diakroninen vai synkroninen. Tämän tutkimuksen aineistolähteessä on ainoastaan uusia yhdyssanoja, joiden asua aika ei ole vielä ehtinyt muuttaa.

2.1.3 Suhde sanaliittoihin

Yhdyssanojen ja sanaliittojen erottamista koskeva keskustelu heijastelee yhdyssanoihin liittyviä laajempia kysymyksiä siitä, kuuluvatko yhdyssanat sanastoon vai kielioppiin ja ovatko niiden muodostusprosessit ensisijaisesti morfologisia vai syntaktisia operaatioita (ks. esim. Lieber 1992; Mäkisalo 2000b: 577; Olsen 2001; Grandi 2009). Käsittelen sanaston ja kieliopin rajankäyntiä luvussa 2.3.1; tässä luvussa keskityn ainoastaan prototyyppisiä yhdyssanoja ja sanaliittoja erottaviin seikkoihin.

Yhdyssanojen painotus vaihtelee eri kielissä. Kuitenkin monissa kielissä – suomi mukaan lukien – painotusta puheessa on pidetty yhtenä tärkeänä yhdyssanan ja sanaliiton erottamiskeinona (Lieber–Štekauer 2009: 8–11; suomen osalta mm. Penttilä 1963 [1957]: § 222; Erkki Itkonen 1966: 237–238; Vesikansa 1989: 213; Jussi Niemi 2009: 239). Suomen kielessä yhdyssanan pääpaino on yleensä koko muodosteen ensimmäisellä tavulla, kun taas sanaliitoissa painotetaan kaikkien osien ensimmäistä tavua. Vaikka sääntöön on poikkeuksia (esim. Penttilä 1963 [1957]: § 222; Vesikansa 1989: 213; Häkkinen 1990: 147), painotusta voitaneen pitää yhtenä prototyyppisen yhdyssanan kriteerinä. On kuitenkin hyvä huomata myös se, että monet eri kieliä tarkastelleet tutkijat ovat myös kiistäneet painokriteerin todistusvoiman sekä yhdyssanojen tunnistamisessa (saksan osalta esim. jo Paul 1898: 304,

15 *-tar*-johtimen tosin oletetaan kehittyneen itsenäisestä *tytär*-sanasta (L. Hakulinen 1979: 182).

englannin osalta mm. Giegerich 2009) että ylipäänsä sanojen erottamisessa puheen virrasta (suomen osalta mm. Suomi 1985: 218–219). Erityisesti yhdyssanoihin liitettyinä painokriteerin ovat suomen osalta kiistäneet Nieminen (1996) sekä Nieminen–Lantee (2011: 72). Heidän mukaansa yhdyssanaa voidaan käytännössä painottaa kuten mitä tahansa sanajonoa: joko yhden tai useamman pääpainon kanssa tai täysin ilman pääpainoa. Myös Fred Karlsson (1983: 150–151) on huomauttanut, että joissakin tapauksissa saatetaan painottaa myös sellaisia elementtejä, jotka eivät ole varsinaisia sanoja (esim. kaksiosaisiksi hahmottuvien lainojen painotus, mm. *árkkitehti*).

Painotus on osaltaan vaikuttanut suomen kielen yhdyssanojen oikeinkirjoitus-sääntöjen muotoutumiseen. Useissa oikeinkirjoitusohjeita antavissa teoksissa kehoitetaan kuulostelemaan sananjonon painotusta oikean kirjoitusasun valitsemiseksi (esim. Ikola 1968: 210; T. Itkonen 2007: 28; KOK 2007: 244). Vaikka painotus on sinänsä irrelevantti kriteeri tässä kirjoitettua kieltä tarkastelevassa tutkimuksessa, oikeinkirjoituksen kautta se kuitenkin vaikuttaa aineistonrajaukseeni (ortografisista yhdyssanoista tarkemmin luvussa 2.2.1; aineistonrajauksesta tarkemmin luvussa 4.2).

Painotuksen lisäksi toinen perinteisesti yhdyssanoja sanaliitoista erottavana pidetty piirre on kokonaisuuden eriytyminen osistaan. Eriytyminen voi olla äännteellistä tai semanttista (myös painotusta voidaan pitää yhtenä äännteellisen eriyty-misen tyyppinä). Sama yhdyssana voi olla sekä äännteellisesti että semanttisesti osistaan eriytynyt, mutta aina semanttinen ja äännteellinen eriytyminen eivät kulje käsi kädessä. (Paul 1898: 301–305; Erkki Itkonen 1966: 237; Bauer 1978: 76.) Yhdyssanojen äännteellinen eriytyminen tapahtuu suomen kielessä käytännössä vakiintuneen ja jatkuvassa käytössä olevan yhdyssanan kuluessa äännteellisesti (esim. *sen tähden* > *sentähden* > *sentään*, Erkki Itkonen 1966: 239–240)¹⁶. Yhdyssanojen äännteellinen eriytyminen osistaan on suomessa hyvin harvinaista, eikä sitä nähdäkseeni voida pitää prototyyppisen yhdyssanan ominaisuutena suomen kielessä. Sen sijaan semanttinen osistaan eriytyminen sopii hyvin kuvaamaan prototyyppistä yhdyssanaa.

Aiemmassa tutkimuksessa yhdyssanojen tärkeänä ominaisuutena on pidetty niiden tavalla tai toisella erikoistunutta merkitystä verrattuna vastaavaan sanaliittoon (esim. Paul 1898: 301–305; Erkki Itkonen 1966: 236–237; Vesikansa 1989: 216; Häkkinen 1990: 146). Periaatteessa ero semanttisesti osistaan eriytyneen ja eriytymättömän muodosteen välille voidaan tehdä helposti¹⁷. Klassinen esimerkki on

16 Osaa yhdyssanojen morfologisista erityispiirteistä on pidetty myös äännteellisinä. Esim. sellaisissa sanoissa kuin *herras mies* ja *pappissääty* esiintyvä linkityselementti *s* voidaan katsoa joko morfologiseksi tai äännteelliseksi. Joissakin tapauksissa *s*-elementti saattaa olla tullut sanaan väärin segmentoidun lainasanana mukana (esim. *saarna|tuoli* vrt. ru *stol*). Monissa tapauksissa elementille ei kuitenkaan ole muuta selitystä kuin aiempien muodosteiden (sekä väärin segmentoitujen lainasanojen myötä syntyneiden että konsonanttivartaloalkuisten *hevosmies*-tyyppiä olevien sanojen) tarjoama analoginen malli. (Häkkinen 1990: 154–155.)

17 Näin siitä huolimatta, että kielellisten ilmausten merkitys on yleensä aina jotakin muuta kuin niiden osien tarkka summa (Fried–Östman 2005).

ilmaisupari *talon poika* ja *talonpoika*, joista ensimmäinen viittaa tietyn talon poikaan eikä ole semanttisesti eriytynyt osistaan, jälkimmäinen taas tietyn ammatin harjoittajaan ja on semanttisesti eriytynyt osistaan. Semanttisesti osistaan eriytyneissä muodosteissa myös kokonaisuuden käsitteellinen yhteenkuuluvuus eli kiteytyneisyys (tai termimäisyys, ks. esim. Vesikansa 1989: 216) on selvää. Käsitteellinen yhteenkuuluvuus tai termimäisyys ei kuitenkaan edellytä kokonaisuuden semanttista eriytymistä osistaan (Vesikansa mp.). *Talonpoika*-esimerkin avulla kuvataan usein myös sitä, että yhdyssanan määriteosa on aina merkitykseltään yleinen, eikä määriteosa (*talon*) voi siksi viitata ilmauksessa mihinkään tiettyyn taloon (esim. Erkki Itkonen 1966: 237).

Kuten todettua, käytännön valintatilanteissa osuvakaan prototyypin kuvaus ei välttämättä ole riittävä keino kategorioiden rajaamiseen. Tämä näkyy myös suomen ortografiassa, jossa yhdyssanan osat kirjoitetaan perussäännön mukaan yhteen mutta jossa kuitenkin monet ilmaukset voivat esiintyä joko yhteen tai erikseen kirjoitettuina merkityksen juurikaan muuttumatta (Erkki Itkonen 1966: 236; ISK § 402). Ilmauksen käsitteellisen yhteenkuuluvuuden hahmottaminen voi vaihdella myös kielenkäyttäjittäin (Niinimäki 1991: 203–205, 1992). Erilainen hahmotus voi johtua esimerkiksi erilaisesta maailmantiedosta (esim. *mustavyö* voi olla kamppailulajien harrastajalle kiinteämpi merkityskokonaisuus kuin niitä harrastamattomalle) mutta myös siitä, että kieleen vakiintumassa oleva ilmaus elää parhaillaan jonkinlaista tiivistymiskautta ja on siten välitilassa myös käsitteellisen yhteenkuuluvuuden suhteen (Niinimäki 1991: 253–254).

Prototyypisten yhdyssanojen lisäksi myös jotkin sanaliitot ovat termimäisiä. Esimerkiksi erikoiskielten sanakirjoissa myös sanaliitot voivat olla määriteltyjä termejä ja esiintyä samanarvoisina käsitteiden niminä kuin yhdistämättömät sanat ja yhdyssanatkin (Haarala 1981: 17). Yleiskielen sanakirjoissa sanaliitot esitellään yleensä jommankumman osansa kohdalla sanan erikoiskäytöksi määriteltynä. Kuitenkin monet sanaliitot ovat myös yleiskielessä yhden yhtenäisen käsitteen nimityksiä samaan tapaan kuin yhdyssanatkin. Erityisesti sellaisissa sanaluokissa, joissa yhdyssanojen muodostaminen on jollain tapaa rajallista, sanaliitot tiivistyvät usein termimäisiksi kokonaisuuksiksi myös yleiskielessä (esim. liittoverbit *pistää silmään*, *ottaa vaarin* tai liittopartikkelit *muun muassa*) (Häkkinen 1990: 147–148).

Yhdyssanojen käsitteelliseen yhteenkuuluvuuteen sekä semanttiseen osista eriytymiseen liittyy kiinteästi ajatus yhdyssanojen leksikaalistumisesta. Leksikaalistumiselle on esitetty monenlaisia määritelmiä, jotka voidaan jakaa karkeasti kahteen ryhmään: suppeisiin ja laajoihin. Suppeiden määritelmien mukaan leksikaalistuminen on monimorfeemisen sanan vakiintumista jakamattomaksi kokonaisuudeksi tiettyyn merkitykseen siten, että sen osat käyvät samalla läpi joko äänteellisiä, morfologisia tai semanttisia muutoksia. Laajojen määritelmien mukaan puolestaan leksikaalistuneita ovat kaikki tiettyyn merkitykseen vakiintuneet sanat. (Koski 1981.) Monet tutkijat katsovat nimenomaan leksikaalistumisen olevan se seikka, joka liittyy yhdyssanat sanastoon ja joka näin ollen on osoitus niiden morfologisesta luonteesta (Bauer 1988: 101–102, 1998: 67–68, Spencer 1991: 312). Tosin myös monet

sanaliitoista voidaan katsoa leksikaalistuneiksi ilmauksiksi (esim. Nenonen 2002: 25–32; Booij 2005: 188).

Suppeaa leksikaalistumisen määritelmää kannattavien tutkijoiden mukaan merkityksen muutos on keskeinen leksikaalistumiseen liittyvä kriteeri, joten yhdysanoista leksikaalistuneiksi voidaan katsoa vain ne, jotka ovat eriytyneet osistaan semanttisesti ja siten menettäneet läpinäkyvyytensä (esim. Bauer 1978: 76; Koski 1981: 19). Varsinkin transformaatioteoriaan pohjaavien näkemysten mukaan osistaan semanttisesti eriytymättömät yhdyssanat ovat leksikaalistumattomia, koska ne generoidaan produktiivisten sääntöjen mukaan ja ne pystytään myös tulkitsemaan pelkän rakenteensa nojalla (ks. esim. A. Hakulinen – Karlsson 1979: 47). Toisin sanoen niiden ”merkitys on niiden osien summa”. Esimerkiksi Auli Hakulisen ja Fred Karlssonin (mp.) mukaan tällaisia yhdyssanoja ovat mm. *lyijykynä*, *vaatekaappi*, *vuorenhuippu*. Mäkisalo (2000b: 578) esittää, että yhdyssanojen semanttinen läpinäkyvyys on yhteydessä niiden osien sanaluokkaan. Hänen mukaansa kahden nominin (esim. *kivitalo*) välinen merkityssuhde ei ole pääteltävissä osien perusteella vaan sanan ymmärtäminen vaatii tietoa maailmasta. Sen sijaan deverbaali jälkiosa osoittaa osien välillä olevan merkityssuhteen olevan sama, kun vastaavan verbin sisältävässä lauseessa (esim. *autonkuljettaja ~ hän kuljettaa autoa*). Funktionaalisen yhdyssanateorian mukaan yhdyssanan merkitys määräytyy aina vasta kontekstissa eikä pelkän yhdyssanan perusteella voida päätellä muodosteen merkityksestä muuta, kuin että sen osilla on merkityksessä jotain tekemistä toistensa kanssa (tarkemmin luku 2.3.1). Funktionaalisen teorian mukaan yhdyssanoja voidaan siis pitää vain morfologisesti läpinäkyvinä. Tämä tutkimus pohjaa funktionalistiseen yhdyssanateoriaan, mutta muiden yhdyssanateorioiden esittelyn yhteydessä viitataan termillä *läpinäkyvä* kuitenkin myös ns. semanttisesti läpinäkyviin yhdyssanoihin.

Toiset tutkijat huomauttavat, että yhdyssanat voivat vakiintua kielenkäyttöön myös silloin, kun ne eivät ole selvästi eriytyneet osistaan semanttisesti. Tämän vuoksi he erottavat leksikaalistuneiden ja pelkän rakenteensa nojalla tulkittavissa olevien yhdyssanojen sijaan kolme vakiintuneisuudeltaan erityyppistä yhdyssanojen ryhmää. (Kosonen 1985: 30–31; Niinimäki 1991: 17.) Esimerkiksi Kirsti Kosonen (1985: 30–31, 89) käyttää termejä leksikaalistuneet, vakiintuneet ja tilapäiset yhdyssanat, joista leksikaalistuneet ovat osistaan semanttisesti eriytyneitä, vakiintuneet sanastoon pysyvästi juurtuneita mutta rakenteeltaan läpinäkyviä ja tilapäiset vakiintumattomia sekä rakenteeltaan läpinäkyviä (ks. myös esim. Bauer 1978: 77). Mäkisaloon (2000a: 14) tutkimuksen mukaan jopa 58 % suomen kielen yhdyssanamuotoisista ilmauksista on kertamuodosteita eli tilapäisiä yhdyssanoja¹⁸. Ei olekaan selvää, että prototyypinen suomen kielen yhdyssana olisi leksikaalistunut tai edes vakiintunut (tosin prototyypisinä pidetyt tapaukset eivät välttämättä aina ole yleisimpiä, ks. esim. Paulsen 2011: 345–365).

18 Helsingin Sanomien korpuksen pitkiä yhdyssubstantiiveja tutkineen Anna Lanteen (2010: 26) aineistosta 77,5 % ja saman korpuksen verbaaliyhdyssanoja tutkineen Suvi Levosen (2006: 130–131) aineistosta 59 % on kertamuodosteita.

Sekä yhdyssanan eriytyminen sen osien merkityksestä, muodosteen käsitteellinen yhteenkuuluvuus eli kiteytyneisyys että leksikaalistuminen ja vakiintuminen ovat asteittaisia ilmiöitä, jotka liittyvät pääasiassa semantiikkaan ja pragmatiikkaan. Esimerkiksi Matthews (1991: 95–96), Booij (2007: 188) sekä Lieber ja Štekauer (2009: 5) huomauttavatkin, että yhdyssanojen määrittelyssä semanttiset kriteerit eivät riitä, vaan hyvään määritelmään tarvitaan myös muodollisia kriteerejä.

Yhtenä yhdyssanoja sanaliitoista erottavana ominaisuutena on pidetty kokonaisuuden morfologista eheyttä ja koskemattomuutta, joka yksittäisten muodosteiden leksikaalistumisen tai vakiintuneisuuden asteesta riippumatta kielii yhdyssanojen ryhmän leksikaalisesta luonteesta (vrt. sanan kriteerit, esim. Karlsson 1994: 75–79). Morfologisen eheyden toisiinsa liittyviä eri puolia ovat 1) yhdysosien pysyvä järjestys, 2) alkuosaan viittaamisen eli anaforan mahdottomuus, 3) sisäisen taivutuksen mahdottomuus sekä 4) se, että yhdyssanan osana voivat esiintyä myös sellaiset muodot, jotka eivät olisi syntaktisessa rakenteessa mahdollisia. Morfologisen eheyden puolista kaksi ensin mainittua kietoutuvat vahvasti myös muodosteen semantiikkaan. Käyn seuraavaksi läpi kaikki neljä yhdyssanojen morfologisen eheyden neljä puolta ja pohdin niiden suhdetta sanaliittoihin.

Varsinkin sellaisten kielten (erityisesti englannin) tutkimuksessa, joissa on vain vähän taivutusta, tärkeimpänä osoituksena yhdyssanan morfologisesta eheydestä nähdään usein osien pysyvä järjestys. Varsinkin generativistisessa perinteessä osien pysyvän järjestyksen on ajateltu olevan seurausta muodosteen jälkiosan pääjäsenyydestä eli siitä, että määrittelyyhdyssanassa toinen osa on semanttisesti hallitseva¹⁹. Tämän vuoksi osien järjestystä ei ole mahdollista muuttaa koko muodosteen merkityksen muuttumatta (esimerkiksi *autokauppa* viittaa eri asiaan kuin *kauppa-auto*). (Spencer 1991: 313; Mäkisalo 2000a: 82; esimerkki Mäkisaloon.) Osien pysyvää järjestystä voidaan pitää osoituksena pitkien ketjuyhdyssanoiksi kutsuttujen muodosteiden yhdyssanuudesta (Scalise–Guevara 2005: 178–179; ks. kuitenkin myös Mäkisalo 2000a: 16–17) (suom. esim. *pyydä ja päästä -kalastaja* [TT]). Myös tässä tutkimuksessa ketjuyhdyssanat on katsottu yhdyssanoiksi ja otettu mukaan tutkimusaineistoon. Niiden poikkeuksellisen rakenteen vuoksi ne on kuitenkin jätetty joidenkin analyysiosassa käytettyjen osa-aineistojen ulkopuolelle (osa-aineistoista tarkemmin luvussa 4.3; lisää ketjuyhdyssanoista esim. Meibauer 2007).

Vaikka osien järjestys on prototyypisessä yhdyssanassa pysyvä, monissa marginaalisemmissä yhdyssanatyypeissä osien järjestys voi vaihdella. Yhdistäminen on monissa kielissä leimallisesti nominien muodostuskeino, ja esimerkiksi yhdysverbit ovat useissa kielissä rakenteeltaan niin löyhiä, että voidaan hajottaa osiinsa merkityksen juuri muuttumatta (esim. suomessa *allekirjoittaa ~ kirjoittaa alle*, *ristinnaulita ~ naulita ristiin*) (Vesikansa 1989: 214). Myös joidenkin genetiivialkuisien yhdyssanojen ero vastaavaan sanaliittoon on niin vähäinen, että niiden osien järjestyksen voi ajatella olevan samaan tapaan ainakin jossain määrin valinnainen

19 Tosin ns. eksosentrisissä yhdyssanoissa ei yleensä katsota olevan lainkaan semanttista pääjäsentä, sillä ne eivät ole edusosansa alakäsitteitä (esim. *herkkusuu* ei yleensä viittaa tietynlaiseen suuhun vaan kokonaiseen ihmiseen).

kuin syntaktisissa rakenteissakin (Mäkisalo 2000a: 82). Tosin järjestyksen vaihtaminen tekee yleensä ilmauksen ensimmäisestä osasta geneerisen sijaan spesifin, kuten myös osien erilleen kirjoittaminen (esim. *aasinnahka* ~ *nahka aasin*). Tämän vuoksi osistaan semanttisesti eriytyneissä yhdyssanoissa muodosteen osien paikan vaihtaminen ei ole mahdollista (*kirkonkylä* ≠ *kylä kirkon*). (Mäkisalo mts. 107–114.) Lisäksi on syytä huomata, että osien pysyvä järjestys koskee periaatteessa vain määrittelyyhdyssanoja. Summayhdyssanoissa yhdysosat ovat keskenään tasavertaisia ja osien järjestys periaatteessa vapaa, joskin usein konventionaalistunut (esim. *kuurosokea* ~ *?sokeakuuro*, *ylösalas* ~ *?alasylös*) tai jonkinlaisen suuntaisuuden motiivoima (esim. *viritin-vahvistin* [ensin virittää, sitten vahvistaa], *suomalais-ruotsalainen sanakirja* [suunta lähtökielestä kohdekieleen]) (ISK § 433).

Toinen sanan sisäiseen eheyteen liittyvä piirre on määriteosaan viittaamisen eli anaforan mahdottomuus (Mäkisalo 2000b: 577–578). Ominaisuus erottaa prototyyppiset yhdyssanat syntaktisista rakenteista, sillä syntaksissa anaforat lausekkeen eri osiin ovat mahdollisia (esim. lauseessa *Oravan häntä on sen tuuhein osa* pronomini *se* viittaa *häntä*-edussanan määritteenä toimivaan *oravaan*). Syntaksissa ovat mahdollisia myös lausekkeen alkuosaan viittaavat määritteet (esim. *tosi punainen paita*), kun taas yhdyssanoissa tällainen viittaaminen ei periaatteessa ole mahdollista (esim. **tosi kevytmaito*). Käytännössä yhdyssanan sisään viitataan kuitenkin joskus esimerkiksi leikittelevässä mainoskielessä (*eräs leipälajeistamme on sen leipominen*) (Mäkisalo 2000a: 85–86; esimerkki Mäkisaloon) tai vaikkapa otsikoissa (*Tosi Tylypahka*, HS Kuukausiliite lokakuu 2011, nro 474, s. 20). Vaikka tällaiset viittaukset tuntuvat epäkieliopillisilta, ne eivät kuitenkaan riko informaatorakennetta tai estä ilmauksen ymmärtämistä. Kompleksisen sanan sisään viittaamisen mahdottomuus onkin ainoastaan syntaksin rajoite, ei semantiikan. Semantiikan lisäksi tulkittaa voi yhdyssanoissa ohjata myös jokin morfosyntaktinen seikka. Jos esimerkiksi edusosan luku sulkee pois siihen viittaavan merkityksen, on viittaus nimenomaan alkuosaan helpompi tunnistaa (*Miten tuo hirsiseinä on noin harmaa, vaikka ne on vasta viime kesänä hiottu?*). (Mäkisalo 2000a: 85–86; esimerkki Mäkisaloon.)

Kolmas erityisesti englannin kieltä käsittelevässä tutkimuksessa usein korostettu morfologiseen eheyteen liittyvä piirre on sisäisen taivutuksen mahdottomuus (esim. Lieber–Štekauer 2009: 5). Suomen kielessä yhdyssanan alkuosan tavallisin sijamuoto on nominatiivi, mutta myös muut sijamuodot ovat mahdollisia. Yhdyssanan sisäisen taivutuksen näkökulmasta fennistiikassa pidetäänkin yleensä poikkeustapauksina vain sellaisia muodosteita, joissa sekä alku- että jälkiosa kongruoivat muun lauseen vaatimuksesta. Yhdyssanan alkuosana taipuvia sanakirjalekseemejä on suomessa noin sata (esim. *omatunto* : *omantunnon*; *uusivuosi* : *uuttavuotta*). Joskus alkuosan taivuttaminen on valinnaista (esim. *isosisko* : *isoasiskoa* ~ *isosiskoa*). (ISK § 420.) Yhdyssanan alkuosan kongruoiminen on mahdollista vain joissakin sanaluokissa. Yhdysnumeraaleissa kaikki osat taipuvat (esim. *viisisataaviisikymmentäkuusi* : *viidensadanviidenkymmenenkuuden*, ISK § 404). Lisäksi alkuosista mahdollisesti kongruoivia ovat adjektiivit ja pronominit (*jompikumpi* : *jommankumman*). Yhdysnumeraalien ja -pronominien lisäksi ainoastaan yhdyssubstantiivien alkuosat

voivat kongruoida. (ISK § 420.) Yhdysverbien kohdalla alkuosan kongruenssi on mahdotonta jo siksi, että finiittiverbi ei voi esiintyä suomessa yhdyssanan määräiteosana (ISK § 430), eikä yhdysverbin määräiteosa siksi voi seurata pääsanaansa persoona-, tempus- ja modustaivutuksessa²⁰. Uudet adjektiivialkuiset yhdyssanat asetuvat suomen kielessä nykyään yleensä taivutusmuottiin, jossa alkuosa ei kongruoi (*valmispitsa* : *valmispitsan*). Myös jotkin vanhastaan kongruoineet alkuosat ovat alkaneet esiintyä myös kongruoimattomina (*mustamaija* : *mustaanmaijaan* ~ *mustamaijaan*). Varsinkin sanaliitoista tiivistymällä voi kuitenkin edelleen syntyä uusia alkuosaltaan kongruoivia yhdyssanoja. (ISK § 421.)

Kansainvälisessä tutkimuksessa ja fennistiikassakin erityisesti yhdyssanojen oikeinkirjoituksen opetuksessa korostettu yhdyssanojen sisäisen taivutuksen mahdolluuteen liittyvä piirre on se, että mikään elementti, esimerkiksi liitepartikkeli, ei voi tunkeutua yhdyssanan osien väliin (esim. Eronen–Maamies–Räikkälä 1996: 6; Mäkisalo 2000a: 82). Kuitenkin eräissä tapauksissa yhdyssanan saamat sidonnaiset morfeemit voivat kiinnittyä muuhunkin kuin jälkiosaan: Osassa *-(i)nen*-loppuisia yhdysadjektiiveja voivat komparatiivin johdin sekä liitteet *-kin* ja *-kAA*n liittyä joko jälki- tai alkuosaan (esim. *laaja-alaisempi* ~ *laajempialainen*; *toisenlainenkin* ~ *toisenkinlainen*). Yhdysnumeraaleissa liitteet *-kin* ja *-kAA*n liittyvät joko viimeiseen tai ensimmäiseen osaan (*viisitoistakin* ~ *viisikintoista*). (ISK § 404.)

Neljänneksi selvästi morfologisesta kokonaisuudesta on kyse silloin, kun alkuosa esiintyy sellaisessa muodossa, jossa se ei voisi esiintyä syntaktisessa rakenteessa (esim. sanavartalot, ks. ISK § 416). Vaikka prototyyppinen yhdyssana koostuu itsenäisistä lekseemeistä (ks. edellinen luku 2.1.2), muodosteen alkuosana toimivat epäitsenäiset yhdysosamuodot (esim. sanavartalot) ja epäitsenäiset leksikaaliset elementit (esim. prefiksinkaltainen *epä-*) toisaalta nimenomaan osoittavat rakenteen yhdyssanaksi. Myös nominatiivimuotoinen alkuosa on selvä osoitus muodosteen yhdyssanuudesta, sillä yhdyssanat ja sanaliitot ovat suomessa eriytyneet toisistaan rakenteellisesti niin, että taipumaton nominatiivimuotoinen substantiivimääräte ei periaatteessa ole mahdollinen sanaliiton alkuosana, kun taas yhdyssanan määräiteosana se on tavallinen (Häkkinen 1990: 145–146, ks. myös Mäkisalo 2000a: 16). Sanaliitoissa poikkeuksina tulevat kyseeseen lähinnä nimekemääritteet (esim. *akrobaatti Alexandra*) (Viinikainen 2010: 313–321).

20 Suomen sukukielissä esiintyy sanaliiton ja yhdyssanan välimaastossa olevia parisanoiksikin nimitettyjä muodosteita, joissa molemmat osat ovat verbejä ja myös taipuvat verbien tapaan (esim. mordvan *oršams-karsems* 'pukeutua' 'pukeutua-kenkiä' tai komipermjakin *piškyny-užny* 'nukkua kuorsaten' 'nukkua-kuorsata') (Erkki Itkonen 1966: 236; suomen osittain samantyyppisistä koloratiivikonstruktioista ks. Rytönen 1937; Jomppanen 2011).

2.2 Suomen kielen yhdyssanat

Normatiivinen oikeinkirjoituksen näkökulma yhdyssanoihin on Suomessa tutuin niin useimmille maallikoille kuin kielentutkijoillekin. Myös suomalaisissa yhdys-sanatutkimuksissa on ollut tavallista rajata yhdyssanojen kategoria oikeinkirjoituksen kautta asiaa lainkaan eksplisiittisesti pohtimatta (ks. myös Mäkisalo 2000a: 64). Myös tässä tutkimuksessa ortografiaa on käytetty tutkimusaineiston rajauksen apuna (eronteko sanaliittoihin on tehty pääasiassa ortografian kautta; johdokset ja yhdyssanat on erotettu prototyypin yhdyssanan ominaisuuksiin nojaavilla perusteilla, ks. luvut 2.1.2 ja 4.2).

Tämän luvun ensimmäisessä alaluvussa (2.2.1) käsitelen suomen kielen yhdysanojen ortografiaa ja vertaan sitä tutkimusteorian kautta määriteltyyn prototyypiseen yhdysanaan (ks. luku 2.1). Luvussa osoitan, että suomen kielen oikeinkirjoitussäännöt pitävät muutamia poikkeuksia lukuun ottamatta yhtä teoreettisen yhdyssanan määritelmän kanssa. Näin ollen osittain suomen kielen ortografian kautta rajatusta aineistosta saadut tulokset voidaan pienin varauksin yleistää koskemaan myös teoreettisten yhdyssanojen ryhmää. Toisessa alaluvussa (2.2.2) käsitelen yhdyssanojen sanatyypin kehittymistä suomen kieleen. Historiallinen näkökulma auttaa havainnollistamaan sitä, minkälaiset rakenteet ovat mahdollisia suomen nykyisissä yhdyssanoissa ja mistä rakenteet ovat saaneet alkunsa. Kolmannessa alaluvussa (2.2.3) tarkastelen yhdyssanojen muodostamista ja morfologista rakennetta nykyisessä suomen kielessä. Luvussa en kuitenkaan enää palaa jo luvussa 2.1 esiin tulleisiin asioihin. Kolmas alaluku perustuu aiempiin suomen kieltä pääasiassa kieliopin tarjoamien mahdollisuuksien näkökulmasta tarkasteleviin tutkimuksiin. Tämän tutkimuksen analyysiosassa käsitelen aiemmassa tutkimuksessa mainittujen mahdollisuuksien hyödyntämistä todellisessa sananmuodostuksessa. (Kieliopillisesta ja käytännöllisestä produktiivisuudesta tarkemmin luvussa 3.1.2.)

2.2.1 Yhdyssanat teoreettisena ja normatiivisena kategoriana

Suomen kielen yhdyssanojen oikeinkirjoituksen perussäännön mukaan yhdyssanan osat kirjoitetaan kiinni toisiinsa. Useissa muissa kielissä ei tehdä kirjoituksen tasolla eroa yhdyssanojen ja sanaliittojen välille. Myös suomen kielen yhdyssanojen oikeinkirjoitus oli pitkään tässä suhteessa vakiintumatonta. Aiemmin yhdyssanojen osat saatettiin kirjoittaa erilleen (*potaatti kuoppa*), yhdysmerkillä yhdistettynä (*aiän-merkki, posti-mestari*) tai suoraan yhteen (*talonpoika, runolaulu*) (Häkkinen 2003: 35). Käytäntö vastasi pitkälti esimerkiksi nykyistä englannin kielen käytäntöä. Suomen kirjakielen normeja alettiin kirjoittaa auki vasta 1820-luvulla (Rintala 1992: 49). Yhdyssanan oikeinkirjoituskäytännön pääperiaatteen voidaan ajatella olleen vakiintunut viimeistään 1800-luvun lopulta alkaen, sillä esimerkiksi E. N. Setälän (1898: 30–31) mukaan ”Yhdistetyt sanat ovat yhteenkirjoitettavat. – – Selvästi yhdistettyjen sanojen erikseen kirjoittaminen, joka on kovin yleistä huolettomassa ja taitamattomassa kirjoituksessa (etenkin ilmoituksissa, kylteissä), on hyvin

pahaksi oikeinkirjoitusvirheeksi katsottava.” (joidenkin erikoistapausten oikeinkirjoituksesta on tosin keskusteltu paljon pidempään, ks. T. Kolehmainen 2014: 157–175).

Pro gradu -tutkielmassaan suomen kielen yhdyssanojen ortografisten normien kehitystä tarkastelleen Heli Kanervamäen (1996) mukaan 1900-luvulla yleisimpiä perusteluja yhteen ja erikseen kirjoittamiselle ovat olleet painotus, selkeys, esteettisyys ja yhtymän merkitys. Suomen kielen pitkiä sanoja on usein pidetty hankalasti hahmottuvina tai epäesteettisinä, ja tämän vuoksi tulkinnanvaraisissa tapauksissa on yleensä suositeltu erikseen kirjoittamista. (Kanervamäki 1996: tiivistelmä ja 18–20.) Esimerkiksi Lari Kotilainen ja Annukka Varteva (2005: 99) luonnehtivat kielenhuolto-oppaassaan suomen kielen yhdyssanojen nykyisiä oikeinkirjoitussääntöjä moninaisiksi ja osin mielivaltaisiksi. Heidän mukaansa sääntöjen kirjavuus johtuu esimerkiksi siitä, että sääntöjä on muokattu moneen otteeseen aina uudenlaisten tapausten ilmaantuessa (ks. myös T. Kolehmainen 2014: 341–347).

Suomen yhdyssanojen nykyiset oikeinkirjoitussäännöt ovat selkeät sellaisten muodosteiden osalta, joiden alkuosa on joko nominatiivissa oleva substantiivi, yhdysosamuoto tai epäitsenäinen leksikaalinen elementti. Tällaisen alkuosan sisältävät sanayhtymät kirjoitetaan aina yhteen (esim. *sammakkolapsi*, *perhoslapsi*, *etälapsi*)²¹. Nämä yhtymät ovat yhdyssanoja myös yhdyssanojen teoreettisen määritelmän mukaan, sillä ne eivät ole mahdollisia syntaktisia rakenteita (ks. tarkemmin luku 2.1.3). Sekä teoreettisesti että ortografisesti yhdyssanoja ovat myös sellaiset rakenteet, joiden määriteosa on nominatiivissa oleva adjektiivi tai numeraali ja edusosa sellainen adjektiivi, joka ei voi esiintyä ilman määritettä (esim. *hyväsydäminen*, *vapaaehtoinen*, *kaksivuotias*). (ISK § 400.)

Jos yhdyssanan alkuosa on adjektiivi tai jossakin muussa sijassa kuin nominatiivissa, kirjoitusasun valintaan vaikuttavat erityisesti merkityksen kiteytyneisyys²² ja vakiintuneet tavat (ks. esim. ISK § 402; Kotilainen–Varteva 2005: 93–100; KOK 2007: 243–246; T. Kolehmainen 2014: 157–175). Kiteytyneisyys eli käsitteellinen yhteenkuuluvuus on myös yksi teoreettisessa yhdyssanan määritelmässä usein käytetyistä kriteereistä. Kielissä, joissa yhdyssanankaltaisia rakenteita ei kirjoiteta yhteen, ilmauksen kiteytyneisyyden ja yhdyssanuuden arviointi jää tutkijalle. Suomen ortografiassa taas yhteen kirjoittaminen on oikeinkirjoituksen perussäännön (yhdyssanat kirjoitetaan yhteen) tuntevan kielenkäyttäjän oma tapa osoittaa, että hän pitää yhtymän merkitystä kiteytyneenä.

Suurin osa suomen kielen yhdyssanojen oikeinkirjoitussäännöistä on säännönmukaisia ja teoreettisten yhdyssanan määritelmien kanssa yhtäpitäviä tai samaan

21 Poikkeus tästä säännöstä ovat substantiivilausekkeen nimekemääritteen ja nimen yhtymät, esim. *prinsessa Ruusunen* (ks. Viinikainen 2010; ISK § 600). Jos tällaisesta ilmauksesta muodostetaan yleisnimi, muodosteen osat kirjoitetaan yhteen (*kaiken maailman prinsessaruususet*). Ks. tarkemmin luku 2.1.1.

22 Kiteytyneisyysääntöön on kaksi poikkeusta: kielten nimet (*suomen kieli*) ja oppiartot (*filosofian maisteri*) kirjoitetaan aina erikseen (Kotilainen–Varteva 2005: 94, 96; KOK 2007: 248; kielten nimiä koskevan normin vakiintumisesta ks. T. Kolehmainen 2014: 165–175). Tällaisia muodosteita on kielessä rajallinen määrä, eivätkä ne todennäköisesti juuri vaikuta suuresta aineistosta saataviin tilastollisiin tuloksiin.

tapaan tulkinnanvaraisia kuin yhdyssanalle esitetyt teoreettiset määritelmätkin. Vain harvat säännöt ovat ristiriidassa yhdyssanan teoreettisten määritelmien kanssa. Eroja on lähinnä sellaisten sanaryhmien ja yksittäisten sanojen kohdalla, joissa oikeinkirjoitussäännöt suosittavat enemmän erikseen kuin yhteen kirjoittamista, vaikka kyseessä olisikin teoreettisesti yhdyssanan kaltainen rakenne. Tällaisia ovat esimerkiksi taipumattomiin tai vajaasti taipuviin sanaluokkiin (eli adverbeihin, adpositioihin ja partikkeleihin) kuuluvat sanat, joiden oikeinkirjoitusta koskevat säännöt ovat usein sanakohtaisia. Näihin sanaluokkiin syntyy yleensä vähemmän uutta sanastoa kuin taipuviin sanaluokkiin, joten produktiivisuutta käsittelevässä tutkimuksessa ne asettuvat joka tapauksessa marginaaliin. Teoreettisesta ja normatiivisesta näkökulmasta eri kategorioihin kuuluvat myös verbin nominaalimuodon ja sen määritteen muodostamat yhtymät (esim. *aikaa vievä, lukuun ottamatta*), jotka suositetaan yleensä kirjoittamaan erilleen (mm. Kotilainen–Varteva 2005: 97; KOK 2007: 252) siitä huolimatta, että niiden on havaittu useiden kielenkäyttäjien sanatajussa hahmottuvan yhdyssanoiksi (Niinimäki 1991: 174–176). Jos yhtymän merkitys on selvästi eriytynyt sen osien merkityksestä, osat kuitenkin kirjoitetaan usein yhteen. Tämä koskee erityisesti muodosteita, joiden jälkiosa on VA-partisiippi (*viinaanmenevä, läpituokeva*, myös esim. *ennennäkemätön*). Sellaisiin yhtymiin, joissa jälkimmäinen osa on verbikantainen substantiivi, pätee oma sääntönsä. Sen mukaan *pano-* ja *pesu-*tyypin sanat (*tilillepano, pyykinpesu*) kirjoitetaan yleensä yhteen määritteensä kanssa kun taas *paneminen-* ja *peseminen-*tyypin sanat (*tilille paneminen, pyykin peseminen*) kirjoitetaan erilleen määritteestään. (Mm. Kotilainen–Varteva 2005: 97; KOK 2007: 251–253.) Säännön voidaan nähdä heijastavan sitä, että yleensä *tilillepano-*tyypin muodosteet ovat merkitykseltään *tilille paneminen* -tyypin yhtymiä kiteytyneempiä.

Yhdyssanojen oikeinkirjoitustaitojen on jo jonkin aikaa väitetty olevan rapistumassa lähinnä englannin kielen vaikutuksen vuoksi (esim. P. Leino 1992: 174). Esimerkiksi Anneli Niinimäki (1991: 250–251) on todennut peruskoulunsa päättävien yhdyssanataitoja käsittelevässä tutkimuksessaan, että koululaisille vaikeimpia tapauksia ovat pitkät yhdyssanat, vierasperäisen aineksen sisältävät yhdyssanat sekä muussa sijassa kuin nominatiivissa tai genetiivissä olevan alkuosan sisältävät yhdyssanat. Lisäksi yhdyssanaksi hahmottamista vaikeuttaa myös alkuosan konsonanttiloppuisuus. Niinimäen tutkimuksessa myös perustyyppisistä nominatiivialkuisista yhdyssanoista jopa lähes 10 % kirjoitettiin väärin erikseen. Oikeinkirjoitussääntöjen noudattamista painetussa mediassa ei ole tutkittu yhdyssanojen näkökulmasta. Toimittajan ammattitaitoon kuitenkin kuuluu oikeinkirjoituksen osaaminen, ja lisäksi toimittajilla on käytössään tarvittavat normatiiviset apuvälineet sekä oikoluokohjelmat. Havaintojeni mukaan sanoma- ja aikakauslehtien toimitetuissa osissa – joiden kieli tämän tutkimuksen kohteena on – noudatetaan yhdyssanojen oikeinkirjoitussääntöjä edelleen hyvin.

Teoreettisesta näkökulmasta yhdyssanojen määrittely ortografian kautta on ongelmallista. Koska kirjoitettu kieli on sekundaaria puhuttuun kieleen nähden, teoreettisen yhdyssanan määritelmän tulee perustua ortografian sijaan myös puheesta

ja kielen rakenteesta havaittavissa oleviin seikkoihin (esim. Lieber–Štekauer 2009: 8; suomen murteiden yhdyssanoja on käsitelty esim. Tuomi 1989: 28–29; puhutun ja kirjoitetun kielen yhdyssanoja ovat vertailleet Pajunen–Palomäki 1984: 64–65). Kirjoitettua kieltä ja kielen käyttöä koskevassa tutkimuksessa ortografian kautta tehtävällä yhdyssanan määrittelyllä ja aineiston rajauksella on kuitenkin kiistattomia etuja.

Kuten luvussa 2.1 kävi ilmi, yhdyssanojen kategorialle ei pystytty esittämään tarkkarajaista teoreettista määritelmää. Koska kielenhuollon asettamat normit eivät tarvitse teoriaa perusteluikseen (esim. T. Kolehmainen 2014: 165–166; teoreettisten yhdyssanojen kategorian ja suomen yhdyssanojen oikeinkirjoitussääntöjen vastaamattomuutta ovat tosin kritisoineet ainakin Niinimäki 1991, 1992 ja Nieminen–Lantee 2011), niistä on mahdollista rakentaa teoriaa yksiselitteisempiä. Vaikka myös suomen kielen ortografiassa on jonkin verran horjuntaa, ortografian käyttö aineiston rajauksen perustana lisää myös tutkimuksen yksiselitteisyyttä ja tekee tuloksista helpommin tulkittavia.

Kielissä, joissa tutkija ei voi käyttää ortografiaa apunaan, tutkimusaineiston rajaus perustuu aina pitkälti tutkijan henkilökohtaisiin tulkintoihin. Sen sijaan suomen kielen ortografisista kriteereistä on ainakin kaikilla painettuun mediaan toimitettua tekstiä kirjoittavilla samankaltainen tieto, joka perustuu erityisesti yhtenäiseen opetussuunnitelmaan. Yhteen ja erikseen kirjoittamisen kannalta valinnaisissa tapauksissa ortografia auttaa tutkijaa näkemään, millaisia kokonaisuuksia kielenkäyttäjät itse pitävät merkitykseltään niin kiteytyneinä, että kirjoittavat ne yhdyssanoiksi. Ortografian käyttö yhdyssana-aineiston rajaamisessa eroaa tässä suhteessa esimerkiksi englantia koskevissa teoreettisissa tutkimuksissa yleensä käytetyistä rajauksista siis vain siinä, kenen näkökulmasta kiteytyneisyyttä tarkastellaan: tutkijan vai kielenkäyttäjän. Kielen käyttöä tarkastelevassa tutkimuksessa nimenomaan kielenkäyttäjien oma näkökulma on tietysti näistä vaihtoehdoista parempi.

On kuitenkin huomattava, että ortografiaa tutkimusaineiston rajauskriteerinä hyödyntävän tutkimuksen (tämän tai minkään muunkaan) tuloksia ei voida suoraan yleistää koskemaan kaikkia teoreettisen määritelmän mukaisia yhdyssanoja tai esimerkiksi puhuttua kieltä. Kunhan teoreettisten ja normatiivisten yhdyssanojen erot otetaan huomioon, tulosten perusteella voidaan kuitenkin tehdä valistuneita arvioita myös näistä jonkin verran toisenlaisista populaatioista.

2.2.2 Sanatyypin synty ja kehitys

Vanhan ja diakronisessa tutkimuksessa yleisesti hyväksytyyn käsityksen mukaan yhdyssanat ovat sanatyypinä syntyneet tiivistymällä sanaliitoista (Erkki Itkonen 1966: 235; Kastovsky 2009: 328–331; tosin esim. Jackendoff 2009: 111–114 esittää, että sanatyypin voisi olla perua protokielelliseltä ajalta, jolloin kielissä ei vielä edes ollut varsinaista syntaksia). Teorian mukaan suomalais-ugrilaisissa kielissä vanhinta yhdyssanatyyppeä edustavat samanarvoisten lauseenjäsenten sidesanattomasta rinnastuksesta syntyneet summayhdyssanat (esim. *maailma*; Erkki Itkonen 1966: 235; Ravila 1961: 76; indoeurooppalaisten kielten osalta aihetta on käsitelty mm. Kastovsky 2009: 328–331).

Toisin kuin sanatyypin alkuaikoina, suurin osa nykykielen yhdyssanoista ei ole summayhdyssanoja vaan määritysyhdyssanoja (esim. *kivitalo*, joka on edusosansa alakäsite). Määritysyhdyssanojen tyyppin syntyajankohdasta on esitetty erilaisia näkemyksiä. Erkki Itkosen (1966: 238) mukaan määritysyhdyssanojen sanatyyppi on syntynyt suhteellisen myöhään, sillä suomalais-ugrilaisissa kielissä nominatiivimuotoinen substantiivimääräte esiintyy ainoastaan yhdyssanan alkuosana vain itämerensuomalaisissa kielissä sekä saamessa (suomessa lisäksi nimekemäärätteenä, ks. Viinikainen 2010). Muissa suomalais-ugrilaisissa kielissä vastaavat rakenteet voidaan edelleen ymmärtää sanaliitoiksi. Sen sijaan Paavo Ravilan (1961: 76) mukaan määritysyhdyssanojen sanatyyppi on saanut alkunsa jo paljon aiemmin, ennen kuin sijamuodot ovat kehittyneet suomalais-ugrilaisiin kielisiin.

Sekä Erkki Itkosen (1966: 238–239) että Ravilan (1961: 76–77) mukaan määritysyhdyssanoista vanhinta tyyppiä edustavat nominatiivialkuiset muodosteet. Itkonen perustaa näkemyksensä sille, että muut kuin nominatiivialkuiset yhdyssanat ovat itäisissä suomalais-ugrilaisissa kielissä hyvin harvinaisia. Hän tosin toteaa myös, että esimerkiksi mordvassa ja marissa on muutamia sellaisia genetiivialkuisia yhdyssanoja, jotka ovat etymologisesti yhteydessä samaa rakennetta noudattaviin suomalaisiin yhdyssanoihin (esim. mordva *čejereńdol* 'lahopuun loiste' ~ *hiirentuli*, mari *jumõntaga* 'taivaanvuohi' kirj. taivaanpässi). Ravilan (mp.) mukaan määritysyhdyssanojen alkuosa ei ole alkuaan ollut varsinainen nominatiivi, vaan ennemmin eräänlainen epämääräinen sija (*casus indefinitus*), joka on kuitenkin seurannut nominatiivin äänteellistä kehitystä ja alettu vähitellen ymmärtää nominatiiviksi. Tämän vuoksi yhdyssanan määriteosana toimiva nominatiivi voi nykyäänkin olla hyvin monenlaisissa merkityssuhteissa edusosaansa.

Se, että yhdyssanojen ajatellaan sanatyypinä saaneen alkunsa sanaliitoista tiivistymällä, ei tarkoita, että kaikkia yhdyssanoja olisi sanatyypin varhaisvaiheessaan muodostettu ainoastaan näin. Heti sanatyypin synnyttyä uusia yhdyssanoja on voitu alkaa muodostaa myös muilla tavoilla, sillä jo luodut yhdyssanat ovat voineet toimia analogisina malleina uusille muodosteille (Erkki Itkonen 1966: 239; Kastovsky 2009: 329). Uusia yhdyssanoja saadaan kieleen nykyäänkin sanaliitoista tiivistymisen kautta (muodostuskeinoista ks. tarkemmin luku 2.2.3). Ravilan (1961: 76) mukaan sanaliittojen tiivistyminen yhdyssanaksi onkin kaikkina aikoina vaikuttava sananmuodostuskeino. Nykyään yhdyssanat muodostetaan kuitenkin tyypillisesti kahdesta tai useammasta sanasta suoraan yhdistämällä (Vesikansa 1989: 213; ks. myös seuraava luku 2.2.3).

Eri sanaluokkiin kuuluvista yhdyssanoista yhdysverbejä on pidetty suomen kielelle alkujaan vieraina (esim. Rahtu 1984: 409). Suomen kielessä on kuitenkin todennäköisesti ollut jonkin verran yhdysverbejä jo ennen kirjakieltä, sillä Agri-colan teksteissä ja saman ajan käsikirjoituksissa esiintyy useita samoja yhdysverbejä (esim. *lähestulla* [vrt. nyk. *lähestulkoon*], *ylenantaa*). Yhdysverbien vanhimpaan kerrostumaan kuuluu adverbialkuisia verbejä, joita suomen kieleen syntyi varsinkin kirjakielen alkuaikoina. Nykysuomen kauden alkuvaiheessa 1800-luvun

loppupuolella adverbialkuisen tyyppin rinnalle ilmestyi nomini- ja yleensä myös nominatiivialkuisen yhdysverbien tyyppi (esim. *palovakuuttaa*, *valokuvata*). Nuorin kerrostuma on 1900-luvulla alkunsa saanut yhdysadjektiiveista johdettujen yhdysverbien tyyppi (esim. *ajanmukaistaa*, *järkipenäistää*). (Häkkinen 1987: 10–19.) Häkkisen (1987: 19) ja Vaittisen (2003: 47) mukaan kaikki kolme tyyppiä ovat edelleen produktiivisia.

Suurin osa suomen kielessä nykyään käytössä olevista yhdyssanarakenteista on peräisin jo kirjakieltä edeltävältä ajalta. Kuitenkin yhdyssanojen on todettu olevan suomessa yleisiä varsinkin kirjoitetussa kielessä, ja esimerkiksi murteissa niitä esiintyy huomattavasti vähemmän (Tuomi 1989: 28). Tuomo Tuomen (1989: 28) mukaan yhdyssanojen muodostusta on käytetty yleiskielessä tehokkaammin esimerkiksi käsitejärjestelmien kuvaamisessa (esim. *rautatieasema*, *sähkösilitysrauta*). Murteissa puolestaan yhdyssanat ovat pääasiassa kiteytyneitä termejä (esim. *kinnasneula*, *ämmänlänget*) tai erityisen yhdyssanajohtimen avulla muodostettuja (esim. *lehmihaka*, *hevosmies*).

2.2.3 Muodostaminen ja morfologinen rakenne nykysuomessa

Yhdistäminen on ilmeisesti kaikkein yksinkertaisin morfologinen operaatio ja siksi myös laajimmin käytetty sananmuodostuskeino maailman kielissä (esim. Greenberg 1963: 93; Bybee 1985: 106–108). Kielellinen universaali se ei kuitenkaan välttämättä ole (Bauer 2009b: 343–345). Suomen kielessä yhdistäminen on jopa suomelle ominaisena sananmuodostustapana pidettyä johtamista yleisempi keino uuden sanaston luomiseen. Esimerkiksi Kielitoimiston sanakirjan (KS 2004–) vuoden 2012 version uusista sanoista noin 75 % on yhdyssanoja (L. Nuutinen 2012: 7). Leksikkoon vakiintuvan sanaston lisäksi yhdistämällä luodaan myös paljon niin sanottuja tilapäismuodosteita, jotka eivät jää pysyväksi osaksi sanastoa.

Suomen kieleen saadaan uusia yhdyssanoja eri tavoilla. Tyypillisin muodostustapa on suoraan kahdesta (tai useammasta) sanasta yhdistäminen (Vesikansa 1989: 213). Yhdyssanojen muodostaminen on luonteeltaan rekursiivista, eli yhdyssanan osana voidaan käyttää myös aiempia yhdyssanoja (esim. *maatalous* + *oppilaitos* > *maatalousoppilaitos*, ISK § 405). Uusia yhdyssanoja syntyy myös aiemmista yhdysanoista ja sanaliitoista johtamalla (esim. *kansaneläke* > *kansaneläkeläinen* ja *pitkätukka* > *pitkätukkainen*) sekä sanaliitoista tiivistymällä (*harhaan johtava* > *harhaanjohtava*). (Vesikansa 1989: 213, 237–238; ISK § 398–399.)

Yhdyssanoja voi muodostaa suomen kielessä oikeastaan loputtomasti. Kaksi sanaa voidaan yhdistää yhdyssanaksi, jos niiden tarkoitteilla vain on kielenulkoisessa maailmassa jotakin tekemistä toistensa kanssa (ks. myös luku 2.3.1). Koska kielellä kuvattavissa olevan kielenulkoisen maailman rajat ovat siellä, missä ihmisen mielikuvituksenkin rajat, yhdyssanojen muodostusta rajoittavat lähinnä pragmaattiset seikat: se, mistä on kulloinkin tarve puhua. (Häkkinen 1990: 144; Tuomi 1995: 154.)

Yhdyssanojen muotoa koskevia rajoituksia on suomen kielessä vain vähän. Yhdyssanoja kuuluu kaikkiin sanaluokkiin (myös partikkelien ryhmään voidaan

ajatella kuuluvan vähintään yhdyssanamaisia interjektioita [esim. *heipparallaa, uppistakeikkaa*]; tosin partikkelien alaryhmistä ainakin konjunktiot ovat aina morfologiselta rakenteeltaan yksiosaisia [ISK § 813, 857]). Sen sijaan yhdyssanojen alkuosan muodolle on joitakin morfosyntaktisia rajoituksia.

Nominit voivat periaatteessa esiintyä suomen yhdyssanojen alkuosana missä sijassa tahansa. Sijamuodoista yleisimpiä ovat yksikön nominatiivi sekä yksikön ja monikon genetiivi (esim. *kivitalo, auringonnousu, äitienpäivä*, ISK § 409, 411). Muut sijamuodot ovat mahdollisia vain silloin, kun yhdyssanan jälkiosa on deverbaali (esim. *unissakävely, pöytiintarjoilu*, ISK § 418–419; Mäkisalonen [2000b: 578] mukaan deverbaali jälkiosa jopa vaatii alkuosan muuhun kuin nominatiivimuotoon). Partitiivin ja monikon nominatiivin esiintyminen on vielä rajallisempaa: ne voivat esiintyä yhdyssanan alkuosana vain silloin, kun jälkiosa on partisiippi tai sen johdos (esim. *veretseisauttava, henkeäsalpaava*, ks. ISK § 418–419).

Sijamuotojen lisäksi suomen kielen yhdyssanojen alkuosana esiintyy erilaisia epäitsenäisiä yhdysosamuotoja ja epäitsenäisiä leksikaalisia elementtejä. Niitä on jaoteltu muotoryhmiin eri esityksissä eri tavalla (tässä tutkimuksessa käytetyt karkea ja hienojakoinen jaottelu on esitelty analyysiluvussa 5.3). Esimerkiksi Vesikansa (1989: 221–222) erottaa omiksi ryhmikseen sanan vartalat (*ihmiskäsi*) ja erityiset yhdysosamuodot (*aitovieri, enkelparvi, kaivinkäsi, tarttumakäsi*). ISK:ssa (§ 416–417) yhdysosamuodoiksi määritellään sellaiset itsenäisten lekseemien edustumat, jotka eivät kuulu kyseisen sanan taivutusparadigmaan. ISK:ssa siis myös sanavartaloita käsitellään yhdysosamuotoina (ks. myös § 407). ISK:ssa (§ 406) mainitaan lisäksi, että yhdyssanan osana voivat toimia myös sellaiset elementit²³, jotka eivät edusta itsenäistä produktiivisesti taipuvaa lekseemiä. Myös nämä elementit luetaan ISK:ssa yhdysosamuodoiksi (§ 416).

Tässä tutkimuksessa luen yhdysosamuodoiksi vain itsenäistä sanaa edustavat epäitsenäiset muodot (esim. em. *ihmiskäsi*, jonka alkuosa edustaa sanaa *ihminen*). Mitään tiettyä sanaa edustamattomille epäitsenäisille elementeille ei ole fennistiekassa vakiintunutta nimitystä, ja tässä tutkimuksessa nimitän niitä *epäitsenäisiksi leksikaalisiksi elementeiksi* ja lyhyemmin *epäitsenäisiksi elementeiksi* (esim. yhdyssanan *epäitsenäinen* alkuosa *epä-* ei edusta mitään tiettyä itsenäistä sanaa vaikka sitä voidaankin käyttää johdosten kantana).

ISK:n (§ 406) mukaan yhdyssanan osina mahdollisia epäitsenäisiä leksikaalisia elementtejä ovat:

- 1) Taipumattomat tai vajaaparadigmaiset sanavartalat (esim. *epä-lähi-, nykyvalko-, ylä-, äkki-*). ISK:n (§ 406) mukaan tämän ryhmän elementit ovat nominaalisia, mutta tässä tutkimuksessa ne on katsottu sanaluokaltaan epäspesifeiksi. Osa niistä voi toimia myös verbien kantana (esim. *epä- > evätä*).

23 ISK:ssa (§ 406) esitellään ensimmäinen tällaisten elementtien kattava jaottelu. Myös mm. Punttila (1995) ja Tuomi (1995) sivuavat aihetta, mutta keskittyvät ensisijaisesti kielenhuoltoon ja yksittäisten elementtien alkuperään ja käyttöön.

- 2) Vierasperäiset prefiksit (esim. *bi-*, *kilo-*, *mini-* ja *super-*).
- 3) Vain yhdessä tai muutamassa sanassa esiintyvät niin sanotut uniikkivartalot (esim. **kottikärry**, **ventovieras**, **toimenpide**, **punatulkku**; esim. Karlsson [1983: 193–194] nimittää tällaisia yhdyssosia jäännösmorfeemeiksi). Ryhmään kuuluu sekä yhdyssanojen alku- että jälkiosia. Uniikkivartalot tai jäännösmorfit ovat yleensä merkitykseltään hämäriä ja usein alun perin lainojen mukana kieleen saatuja (Karlsson 1993: 194; Liljestrand 1993: 22–23).
- 4) Pelkästään yhdyssanan määräiteosana esiintyvät johdokset (esim. *alkeiskaksois-*, *kohtuu-*, *uusio-* ja *yhtiö-*). Tässä ryhmässä on sekä nominimuotoisia johdoksia (*kohtuu-*, *uusio-* ja *yhtiö-*) että sellaisia *s*-loppuisessa yhdyssanamuodossa olevia alkuosia, joille ei kuitenkaan ole kielessä olemassa lähtömuotoa, josta yhdyssanamuoto olisi muodostettu (esim. *alkeis-* < **alkeinen* ja *kaksois-* < **kaksoinen*).
- 5) Aina seurakseen määräiteosan vaativat ja pelkästään yhdyssanan edusosana esiintyvät johdokset (esim. *-pääsyinen*, *-hoitoinen*, *-mukaistaminen*, *-päiväisyys*, *-vuotias*).
- 6) Suoranaisesti lekseemiä edustamattomat ortografiset elementit (*s-kirjain*, *2,5-kertainen*, *@-merkki*).

ISK:ssa mainittujen tapausten lisäksi uniikkivartaloiden joukkoon voidaan lukea myös erilaiset ekspressiiviset elementit (esim. **örbäturismi** [TT], **mökömökömaa** [TT]). Niistä monet voivat toimia muiden sanavartaloiden tapaan johdosten kantoina, mutta sanakirjafrekvenssinsä puolesta useimmat niistä rinnastuvat ennemmin uniikkivartaloihin.

Yhdyssanan alkuosana toimiville epäitsenäisille leksikaalisille elementeille on yhteistä, että niiden sanaluokkaa ei voida yksiselitteisesti määrittää. Myös itsenäiset lekseemit menettävät syntaktisia ominaisuuksiaan yhdyssanan alkuosana toimiesaan (Spencer 2003: 215), ja myös niiden statusta yhdyssanassa voidaan pohtia. Spencerin (mts. 216) mukaan yhdyssanan osat voivat yhdyssanan leksikaalistuessa menettää myös semanttiset ominaisuutensa. Esimerkiksi englanninkielisen yhdyssanan *blackbird* 'mustarastas' alkuosa *black-* on menettänyt alkuperäisen merkityksensä. Sama esimerkki toimii myös suomenkielisenä: Mustarastaista vain koiraat ovat väritykseltään mustia. Naaraat sen sijaan ovat selkäpuolelta tummanruskeita ja vatsapuolelta vaaleampia. Lajin nimitys *mustarastas* on leksikaalistunut eikä sen alkuosalla *musta-* ei ole enää samaa merkitystä kuin itsenäisellä sanalla *musta*.²⁴ Spencerin (mp.) mukaan tällaisen leksikaalistuneen yhdyssanan sisäinen rakenne heijastaa vain sanan syntyhistoriaa, eikä sen alkuosalla ole enää sen enempää sanaluokkaa kuin itsenäistä merkitystään. Jäljellä on pelkkä morfofonologinen muoto.

²⁴ On tosin otettava huomioon, että *black-* ja *musta-*sanoilla on myös muita merkityksiä kuin pikimusta väri (esim. 'tumma') ja että mustarastaskoiraskaan ei ole kokonaan musta (sen nokka on kirkkaankeltainen). Esimerkki kuitenkin havainnollistaa melko hyvin sitä, että varsinkin yhdyssanan alkuosan merkitys voi sanan leksikaalistuessa muuttua tai alkuosa voi menettää itsenäisen merkityksensä kokonaan.

Koska varsinkin itsenäistä merkitystä pidetään lekseemin tärkeänä ominaisuutena, on kyseenalaista, voidaanko tällä tavoin leksikaalistuneen yhdyssanan osia enää pitää lekseemeinä.

Yhdyssanan osan oman merkityksen täydellinen katoaminen voi koskea vain sellaisia yhdyssanoja, jotka kielenkäyttäjät tulkitsevat aina kokonaisuuden kautta eivätkä koskaan osina (yhdyssanojen prosessoinnista ks. luku 2.3.3). Myös hyvin leksikaalistuneet mutta kuitenkin metaforisen luonteensa säilyttäneet yhdyssanat voidaan todennäköisesti hahmottaa niiden osien kautta (esim. *hiirenkorva*).

Suomen kielessä finiittinen verbi ei voi esiintyä yhdyssanan alkuosana (vrt. sivulla 35 alaviitteessä 20 mainitut sukukielten ns. parisanat), vaan yhdyssanan alkuosassa verbiä edustaa *-mA-* tai *-in-*päätteinen muoto tai infiniittinen verbi-muoto (*juomatapa, keitinliemi, nukkumaanmeno*, ISK § 407, 417). Koska sanalla ei yhdyssanan alkuosana toimiessaan ole syntaktisia ominaisuuksia samassa mielessä kuin itsenäisenä esiintyessään (Spencer 2003: 215–216), verbaalisia merkityksiä voidaan ilmaista yhdyssanan alkuosassa myös nomineilla ja sanavartaloilla (esim. *toteuttamismahdollisuus, seisonta-aika, keruutapa, riippuliito, syntyperä*, ks. myös ISK § 407, 417).

2.3 Yhdyssanojen merkitys ja tulkinta

Perinteisen ja fennistisessä tutkimuksessakin pääasiassa käytetyn tavan mukaan yhdyssanat on jaettu kahteen semanttiseen päätyyppiin, määritysyhdyssanoihin (esim. *kivitalo*, joka on edusosansa alakäsite) ja summayhdyssanoihin (esim. *parturi-kampaamo*, jossa yhdysosat ovat semanttisesti tasa-arvoiset)²⁵, joille on lisäksi erotettu alaryhmiä. Esimerkiksi ISK:ssa (§ 398–431) määritysyhdyssanojen päätöksen alla käsitellään myös appositioyhdyssanat (esim. *tutkijanainen*, kuvaa samaa tarkoitetta eri kannoilta), bahuvriihdyssanat (esim. *herkkusuu*, viittaa tarkoitteeseen sen osan kautta) sekä toistoyhdyssanat (esim. *ruokaruoka*, korostaa tarkoitteen prototyypillisyyttä). Näistä ainoastaan appositioyhdyssanojen mainitaan eksplisiittisesti olevan määritysyhdyssanojen alalaji (mt. § 422), vaikka toisaalta tätä ennen todetaan, että määriteosan voidaan mieltää appositioyhdyssanoissa ”rajaavan tarkoitteiden joukkoa **samaan tapaan** kuin määritysyhdyssanoissa” (korostus omani).

Kaiken kaikkiaan fennistiselle tutkimukselle tuntuu olevan tyypillistä semanttisten yhdyssanatyypien esittelyn luettelomaisuus, käsitejärjestelmän

25 *Määritysyhdyssana* ja *summayhdyssana* ovat ISK:n (2004: 388) käyttöön ottamia termejä. Aiemmin yhdyssanatyypeistä käytetyt nimet vaihtelevat: Esimerkiksi Erkki Itkosen (1966: 234) mukaan yhdyssanat voivat olla joko kopulatiivisia eli rinnasteisia tai determinatiivisia eli alisteisia. Näitä termejä käyttää myös Vesikansa (1989: 213). Penttilä (1963 [1957]: § 223, s. 260–261) puolestaan käyttää nimityksiä *summatiiviset yhdyssanat* ja *determinatiiviset yhdyssanat*. Kolmanneksi pääryhmäksi hän erottaa iteratiivisten yhdyssanojen eli ISK:n termein toistoyhdyssanojen ryhmän. *Kopulatiiviseksi yhdyssanoiksi* Penttilä kutsuu appositiorakenteen sisältäviä yhdyssanoja.

eksplikoimattomuus sekä jaottelussa käytettyjen kriteerien yhteismitattomuus. Bisetto–Scalise (2005, 2009: 35–44) kritisoivat universaaleiksi tarkoitettuja yhdys-sanojen semanttisia luokitteluja osin samoista asioista. Tässä luvussa käsittelen suomen kielen yhdyssanojen merkityksen rakentumista kokonaan ilman perinteisiä semanttisia yhdyssanatyyppejä.

Tämän luvun ensimmäisessä alaluvussa (2.3.1) pohdin kysymystä yhdyssanojen kuulumisesta leksikkoon ja kielioppiin. Luvun luonteva osa ovat myös tärkeimmät yhdyssanojen merkityksen rakentumista koskevat yleiskielitieteelliset teoriat, jotka ottavat kantaa siihen, onko yhdyssanojen merkityksen rakentuminen sanakohtaista vai jonkinlaisten yleisten sääntöjen ohjaamaa. Tämän tutkimuksen kannalta tärkein teoria on funktionalistinen yhdyssanateoria, jonka mukaan yhdyssanan osien välinen merkityssuhde määräytyy aina lopullisesti vasta kontekstissa.

Funktionalistinen yhdyssanateoria sopii parhaiten nominatiivialkuisten yhdys-sanojen merkityksen käsittelyyn. Suomen kielessä, jossa yhdyssanan alkuosa voi olla periaatteessa missä sijamuodossa tahansa, alkuosan muoto voi osoittaa yhdys-sanan osien välisen merkityssuhteen yhtä tarkasti kuin sijamuodot tekevät syntaktisessa rakenteessakin. Esimerkiksi yhdyssanassa *maahanimeyttämö* [TT] osien välinen merkityssuhde on selvä ilman kontekstiäkin. Niinpä suomen kielen yhdys-sanojen merkityksen ja tulkinnan kannalta on tärkeää käsitellä myös niiden alku-osan muodon vaikutusta (toinen alaluku 2.3.2).

Kolmannessa alaluvussa (2.3.3) käsittelen yhdyssanojen prosessointia ja esitte-len aihetta koskevia psykolingvistisiä teorioita ja tutkimuksia, sillä yhdyssanojen prosessointiin ja ymmärrettävyyteen liittyvät seikat vaikuttavat myös yhdyssanara-kenteiden produktiivisuuteen.

2.3.1 Kuuluvatko yhdyssanat leksikkoon vai kielioppiin?

Fennistiikassa yhdyssanoja on kuvattu pääasiassa leksikon osana eli kahden sanan morfologisenä yhdistymänä, joka voidaan muodostaa, kun tietyt semanttis-prag-maattiset ehdot täyttyvät (Mäkisalo 2000a: 68). Kansainvälisessä yhdyssanatutki-muksessa vallitsee kuitenkin erilaisia käsityksiä siitä, ovatko yhdyssanat muiden sanojen tapaan varastossa leksikossa vai tuotetaanko ne jopa jokaisella käyttöker-ralla uudelleen kieliopin (syntaksin tai kielioppiin kuuluvan erillisen sananmuodos-tuskomponentin) sääntöjä noudattaen (ks. myös Mäkisalo 2000b)²⁶. Tässä luvussa luon lyhyen katsauksen aiemmissä tutkimuksissa esitettyihin näkökohtiin. Eniten tilaa saavat maailmalla edelleen suosiossa oleva generativistinen näkökulma sekä tämän tutkimuksen kannalta keskeinen funktionalistinen yhdyssanateoria.

Yhdyssanoja leksikaalisina yksiköinä tarkastelleen strukturalismin jälkeen generativistit toivat yhdyssanojen tutkimukseen suuren muutoksen. Varhaisen

26 Varhaisessa kielitieteessä myös taivutusmorfologia laskettiin osaksi kielioppia, mutta morfolo-gian irtaannuttua itsenäiseksi tutkimusalakseen 1900-luvulla (kielitieteellisessä merkityksessä termiä *morfeemi* käytti ensimmäisenä August Schleicher 1859) sananmuodostus alettiin käsitellä omana ko-konaisuutenaan (esim. Chomsky 1970).

transformaatioteorian mukaan leksikkoon kuuluivat vain yksimorfeemiset sanat. Yhdyssanojen muodostuksen ajateltiin perustuvan niiden generointiin syntaktisista rakenteista kieliopin sääntöjen avulla. Generativistisen tutkimustradition piirissä tilanne muuttui, kun yhä useammat tutkijat alkoivat käsittää myös kompleksiset sanat leksikkoon kuuluviksi ja syntyi ajatus, jonka mukaan (varsinkin epäsäännönmukaista) sananmuodostusta säätelevät sen omat syntaktisista transformaatioista riippumattomat säännöt (alkuun käsiteltiin ensisijaisesti johtamista, Chomsky 1970; Halle 1973; Aronoff 1976; erityisesti yhdyssanoihin kiinnittää huomiota Allen 1978). Näiden uudennosten myötä generativistisen tutkimussuuntauksen piiriin syntyi leksikalismiksi kutsuttu teoreettinen koulukunta (ks. tarkemmin Scalise–Guevara 2005). Erityisesti sananmuodostuksen ja taivutusmorphologian erottavaan niin sanottuun heikkoon leksikalistiseen hypoteesiin perustuvilla teorioilla on yhä kannattajansa (*Weak Lexicalist Hypothesis*, ks. tarkemmin esim. Scalise–Guevara 2005: 170–171). Eräät myöhempien syntaktisten teorioiden kannattajat ovat kuitenkin esittäneet myös, että sananmuodostusta koskevat erilliset säännöt ovat tarpeettomia, koska kaikki on selitettävissä myös pelkkien syntaksin sääntöjen avulla (Baker 1987).

Nykyään generativistienkin keskuudessa yleisin käsitys lienee se, että syntaksi ja morfologia eivät ole kieliopin erillisiä moduuleja, vaan ne ovat vuorovaikutuksessa toistensa kanssa (Spencer 1991: 309; Mäkisalo 2000b: 577; Scalise–Guevara 2005: 182; erilaisia näkemyksiä erityisesti yhdyssanojen osalta esittelee tarkemmin esim. Mäkisalo 2000a: 29–60). Erilaisista näkemyksistä huolimatta generativistisille yhdyssanatuksille on yhteistä pyrkimys yhdyssanojen muodostuksen taustalla oleviksi oletettujen geneeristen sääntöjen ja syvärakenteiden selvittämiseen. Tutkimuksessa pidetään yleensä tärkeänä leksikaalistuneiden sekä tilapäisten ja vielä vakiintumattomien yhdyssanojen erottamista, sillä nimenomaan uusien sanojen muodostuksen ajatellaan olevan säännöllisten ja produktiivisten sääntöjen säätelemää. Leksikaalistuneisiin sanoihin sen sijaan kuuluu paljon sellaisia jopa arkaaisista muodosteista, jotka eivät välttämättä paljasta enää mitään sananmuodostuksen sääntösystemistä. (Lehrer 1996: 63–73; Mäkisalo 2000a: 9.) Syvärakenteisiin perustuva yhdyssanakäsitys sopiikin kuvaamaan korkeintaan niin sanottuja läpinäkyviä yhdyssanoja (läpinäkyvyydestä ks. tarkemmin s. 32 luvussa 2.1.3).

Syvärakenteiden ongelma on se, että todellisessa kielenkäytössä mahdollisia yhdyssanan osien välisiä merkityssuhteita on rajaton määrä. Kun tästä rajattomien mahdollisuuksien joukosta ryhdytään tekemään yleistyksiä, niistä tulee väkisin kovin abstrakteja (suomen osalta ks. esim. Saukkonen 1973). Abstraktien syvärakenteiden käyttö tutkimuksessa on todettu hankalaksi, sillä ihmiset tulkitsevat niitä hyvin eri tavoin. Abstraktiudesta huolimatta kaikkien yhdyssanojen taustalle ei sitä paitsi aina löydy sopivaa syvärakennetta²⁷ (mahdollisten suhteiden sijaan onkin siksi parempi pohtia sitä, mikä ei ole mahdollista, ks. Downing 1977).

²⁷ Luonnollisessa kielessä on niin paljon epäsäännöllisyyttä, että sitä ei voida täydellisesti kuvata formaalisten kielten tapaan matemaattisen tarkasti. Toisin sanoen mikään luonnollisen kielen kielipillinen kuvaus ei riitä kuvaamaan kaikkia kieleen kuuluvia ilmaisuja, vaan aina on poikkeuksia.

Vastavetona generativistisille yhdyssanateorioille syntyi 1970-luvun lopulla funktionalistinen yhdyssanakäsitys. Sen lähtökohtana on kielenkäyttö, ja nähdäkseni se kuvaakin syvärakenteita paremmin sekä todellisessa kielenkäytössä esiintyvien yhdyssanojen semanttisia ominaisuuksia että suhdetta leksikkoon ja sananmuodostukseen (samalla kannalla ovat mm. Seppänen 1981 ja Räisänen 1986). Funktionalistisen teorian etu syvärakennemalleihin nähden on myös se, että sen avulla pystytään käsittelemään kaikkia yhdyssanoja tilapäisistä leksikaalistuneisiin asti (Seppänen 1981: 75–76).

Funktionalistisen yhdyssanateorian perusajatus on, että yhdyssanoilla ei ole kielen systeemin (eli kieliopin) tasolla minkäänlaista syvämerkitystä, vaan ilman kontekstia yhdyssanan merkityksestä voidaan päätellä vain, että sen osilla on koko muodosteen merkityksessä jotakin tekemistä toistensa kanssa²⁸. Funktionalistisen teorian mukaan sanan kulloinenkin merkitys siis määräytyy aina vasta kontekstissa. Generativististen ja funktionalististen teorioiden tärkeimmän eron voidaankin katsoa olevan siinä, mihin ne vetävät kielen ja kielenulkoisen todellisuuden rajan: generativististen teorioiden mukaan yhdyssanojen syvärakenteet ovat osa kieltä, kun taas funktionalististen näkemysten mukaan kyse on oikeastaan kielenulkoisen maailman rakenteista, joita ei voi tulkita pelkän kielitiedon nojalla. (Seppänen 1981.)

Funktionalistinen yhdyssanakäsitys perustuu Eugenio Coseriuun (1971, 1975; ks. myös Seppänen 1981; Pitkänen 2008: 102–104) Ferdinand de Saussuren *langue et parole* -ajatuksesta kehrittelemään malliin, jonka mukaan kielen sanasto rakentuu kolmiportaisesti. Mallissa yhdyssanat ovat olemassa systeemin, normin ja puhunnan (System – Norm – Rede) tasolla, ja sanojen merkitykset rakentuvat ja määräytyvät näillä tasoilla eri tavoin.

Systeemin (eli käytännössä kieliopin) tasolla ovat olemassa kaikki sellaiset yhdyssanat, jotka kielen kielioppi sallii. Sanoja ei välttämättä ole koskaan käytetty, mutta periaatteessa kuka tahansa voisi käyttää niitä milloin tahansa, jos maailmassa vain olisi sopiva tarkoite, johon sanoilla voisi viitata. Suomen kielioppi ei juuri rajoita sitä, mitä sanoja voidaan yhdistää yhdyssanoiksi (esim. englannissa rajoituksia on paljon enemmän, ks. Ryder 1994: 4–6). Yhdyssanojen muodostuksen rajoitteet ovat kielenulkoisessa maailmassa, joten kielen systeemin tasolla on olemassa periaatteessa loputon määrä yhdyssanoja. Systeemin tasolla yhdyssanojen osien välinen merkityssuhde ei ole vielä millään tavalla määräytynyt, vaan kaikki merkitystulkinnat ovat mahdollisia. Esimerkiksi *kalakauppias* voi tarkoittaa mitä tahansa kauppiasta, jolla on jotain tekemistä kalan kanssa (esim. ’kauppias, joka myy kaloja’, ’kala, joka toimii kauppiana’) (Räisänen 1986: 160). Kielen systeemin tasolla yhdyssanojen merkityksen kaikki semanttiset ovet ovat siis vielä auki.

Coseriuun (1971, 1975) mallissa **normiin** (eli vakiintuneiden merkitysten tasoon, leksikkoon) kuuluvat ne ilmaukset, jotka kieliyhteisö on ottanut yhteiseen käyttöön ja joille on siksi kehittynyt yhteisesti hyväksytty merkitys. Kun sana

28 Tosin tämänkin merkitys voidaan tietysti käsittää erittäin abstraktiksi syvärakenteeksi: x on jossain suhteessa y:hyn.

otetaan kieliyhteisön yhteiseen käyttöön, sen merkitys tarkentuu ja muut systeemi-merkityksessä avoinna olleet semanttiset ovet sulkeutuvat. Esimerkiksi *kalakauppiaan* normimerkitys nykyisessä kielessämme on 'kauppias, joka myy kaloja'. Vaikka normi ohjaa yksittäisten puhujien kielenkäyttöä, se ei kuitenkaan voi yksiselitteisesti määrätä sitä. Coseriun mallissa **puhunta** (eli todellinen kielenkäyttö) sisältää yksittäisen kielenkäyttäjän tilannekohtaiset mahdollisuudet käyttää kieltä – myös normin vastaisesti. Puhunnassa vasta konteksti ratkaisee yhdyssanan merkityksen, ja puhunnassa yksittäinen kielenkäyttäjä voi jälleen ryhtyä availemaan normin kertaalleen sulkemia ovia. Sopivassa kontekstissa esimerkiksi *kalakauppias* voi viitata vaikkapa jollakin tapaa kalaa muistuttavaan kauppiaaseen. Sadussa *kalakauppias* voi puolestaan olla myös kauppiaana toimiva kala. Tilannekohtaisessa käytössä sana voi viitata myös johonkuhun tiettyyn kalakauppiaaseen.

Mikäli normiin vakiintunutta yhdyssanaa käytetään sen normimerkityksestä poikkeavasti, tarkoitetun merkityksen on yleensä käytävä jollain tapaa ilmi kontekstista. Myös täysin vakiintumattomien yhdyssanojen merkitys määräytyy viime kädessä vasta kontekstissa. Yhdyssanat ja johdokset poikkeavat kuitenkin uusista yksimorfeemisista sanoista siinä mielessä, että kielenkäyttäjä voi tulkita niiden merkityksen aina jollain tavalla, vaikka kontekstia ei olisikaan saatavilla. Coseriun (1971, 1975) mallin mukaan vaikuttaa siltä, että yhdyssana voi vakiintua mihin tahansa kielen systeemin sallimaan merkitykseen (rajoitteita mainitsee esim. Downing 1977). Yhdyssanojen muodostaminen on kuitenkin niin produktiivinen prosessi, että muodosteiden merkityksen määräytyminen ei voi olla täysin mielivaltaista. Esimerkiksi Mary Ellen Ryderin (1994: 6) mukaan sananmuodostuskeinon produktiivisuus kärsisi väistämättä, mikäli kielenkäyttäjillä ei olisi minkäänlaisia keinoja ennustaa muodosteen merkitystä ilman kontekstia.

Kielenkäyttäjän tekemiä merkitystulkintoja ohjaa ilmauksen mahdolliseen vakiintuneeseen merkitykseen liittyvän tiedon lisäksi myös maailmantieto. Esimerkiksi ilmaus *metsästäjän ampuminen* on monitulkintainen, mutta maailmantieto ohjaa tulkitsemaan sen ensisijaisesti metsästäjän suorittamaksi ampumiseksi (Mustaajoki 2009: 65). Samasta syystä esimerkiksi ruotsin *krokodilväska* osataan tulkita ensisijaisesti krokotiilinnahasta valmistetuksi laukuksi ja *damväska* ensisijaisesti naisten käyttämäksi laukuksi (Räisänen 1986: 158).

Suomen kielen kannalta funktionalistisen yhdyssanateorian ongelma on se, että periaatteessa se ottaa huomioon ainoastaan nominatiivialkuiset kahdesta substantiivista koostuvat yhdyssanat. Suomessa (kuten monissa muissakin kielissä) myös yhdyssanan morfologinen rakenne voi antaa vihjeitä muodosteen merkityksestä. Koska funktionalistinen teoria ei yksinään riitä taustaksi suomen kielen yhdyssanarakenteita käsittelevälle tutkimukselle, sitä on täydennettävä huomioilla alkuosan muodon vaikutuksesta yhdyssanan merkityksen rakentumiseen (käsittelen tätä aihetta seuraavassa luvussa 2.3.2; deverbaalin jälkiosan vaikutuksesta ks. myös Mäkisalo 2000b: 578).

Generativistien tiukkojen sääntöjen ja funktionalistien hyvin abstraktien rakenteiden ja sattumanvaraiselta vaikuttavan vakiintumisen ja käytön lisäksi

yhdyssanojen merkityksen rakentumista on kuvattu myös erilaisten analogioiden synnyttämän todennäköisten merkitysten verkoston kautta. Esimerkiksi Ryderin (1994; ks. myös Malmivaara 2004) mukaan yhdyssanojen muodostus perustuu pääasiassa analogiseen tietoon aiempien yhdyssanojen muodosta, merkityksestä ja käytöstä, ja vain kaikkein produktiivisimmissä tyypeissä analogian voidaan ajatella muuttuvan eräänlaiseksi säännöksi. Ryderin teoreettinen tausta nojaa pitkälti kognitiivisen kieliopin peruskäsitteeseen, skeemaan (ks. erityisesti Langacker 1987, 1988; Rumelhart–Ortony 1977; Mandler 1984). Suoraan aiempaan muodosteeseen perustuvan analogian²⁹ ja abstraktin skeeman (generativistien termein säännön) raja ei ole tiukka, sillä skeemat tai säännöt voidaan nähdä myös riittävän usein toistuneista analogioista muodostuneina abstraktioina (näin on todennut mm. jo Setälä 1898: 114).

Skeemalle läheisiä käsitteitä ovat konstruktio ja muotti (skeemasta yhdyssanojen kuvauksessa Ryder 1994; Heyvaert 2009; konstruktio ja muotin vertailua Koivisto 2006: 38; konstruktioista yhdyssanojen kuvauksessa Booij 2009). Sekä skeema, konstruktio että muotti ovat oikeastaan omalla tavallaan määriteltyjä kieliopillisen säännön vastineita, joita on käytetty eri tutkimussuuntauksissa³⁰. Sääntöihin nähden skeeman, konstruktio ja muotin suurin etu on, että niiden avulla voidaan kuvata sekä säännönmukaiset että idiosynkraattiset tapaukset. Esimerkiksi muotti [*nais-/mies-* + henkilöön viittaava sana] tuottaa yhdyssanoja, joiden tarkoite on 'nais-/miespuolinen henkilö' (esim. *nais* + *urheilija* > *naisurheilija* 'naispuolinen urheilija') (Karlsson 1974, 30; Tyysteri 2011c: 163, 169). Muottiin asetuvat sanat voivat kuitenkin vakiintua myös muihin merkityksiin (esim. *naissankari* 'naistenmies, casanova'), ja todellisessa käytössä merkitys määräytyy aina lopullisesti vasta

29 Analogisena mallina uudelle yhdyssanalle voi toimia mikä tahansa jo olemassa oleva muodoste, joten kaikki kielessä käytössä olevat yhdyssanarakenteet voivat myös tuottaa uusia muodosteita. Analoginen malli uudelle muodosteelle voidaan saada myös vieraasta kielestä, jolloin kyse on käännöslainauksesta (esim. *pilvenpiirtäjä* < en *skyscraper*, L. Hakulinen 1969: 14).

30 Käsitteiden suurin ero on oikeastaan niiden lähtökohdissa (Östman–Fried 2005; Kotilainen 2007: 13–15). Skeeman käsite on lähtöisin alkujaan generativistisen kieliopin vastavoimaksi syntyneestä kognitiivisesta kieliopista, kun taas konstruktio käsitteen alkukotia konstruktiokielioppia on tavattu pitää generativistisena teoriana (ks. esim. Kay 1995: 171; ks. myös Kotilainen 2007: 13–14). Yhteistä konstruktio ja skeeman eri määritelmille on, että molemmat käsitetään yksittäisten ilmausten joukosta tehtäviksi yleistyksiksi, rakennemalleiksi, jotka ovat muodon ja merkityksen yhteenliittymiä (konstruktio määrittelmistä ks. Fillmore–Kay–O'Connor 1988; Östman 1988; Goldberg 1995, 2006; Kotilainen 2007: 13–22; skeeman määrittelmistä ks. esim. Langacker 1987, 1988; Onikki–Rantajääskö 2001: 32). Esimerkiksi Huumo ja Ojutkangas (2005: 223) sekä Jarkko Niemi (2010) käyttävät rinnakkain nimityksiä *konstruktio* ja *muotti*. ISK:ssa (VISK – määritelmät s.v. *konstruktio*) konstruktio määritellään muotiksi tai muusta useammasta osasta koostuvaksi kielen rakenteeksi.

Varsinkin konstruktio käsitettä on käytetty erityisesti perifeeristen syntaktisten ja idiomaattisten rakenteiden kuvaukseen (Suomessa esim. J. Leino 2003; Kotilainen 2007). Yhdyssanojen kuvauksessa näistä kolmesta käsitteestä on ollut käytössä lähinnä skeema (Ryder 1994; Onikki–Rantajääskö 2001: 32; Malmivaara 2004), jota on käytetty yhdyssanojen merkitykseen keskittyvissä tutkimuksissa. Myös konstruktio käsitettä on toisinaan käytetty yhdyssanojen rakenteen kuvaamiseen (mm. Booij 2009). Muotti on ISK:n (2004: 172–173; VISK – määritelmät s.v. *muotti*; ks. myös A. Hakulinen 2006: 341) käyttöön ottama käsite, jota ei ole omaa alustavaa hahmotelmaani (Tyysteri 2011c: 163, 169) lukuun ottamatta toistaiseksi käytetty yhdyssanojen kuvaamisessa.

kontekstissa (esim. *naislääkäri* voi olla naispuolinen lääkäri mutta myös esimerkiksi gynekologi). Muotit tai skeemat eivät siis ole ristiriidassa funktionalistisen yhdyssanateorian kanssa, mutta auttavat ennustamaan mikä kielioopin tasolla mahdollisista yhdyssanan osien välisistä merkitystulkintoista todennäköisesti vakiintuu normin tasolle ja missä merkityksessä yhdyssanaa todennäköisesti käytetään puhunnan tasolla.

2.3.2 Alkuosan muoto ja yhdyssanan merkitys

Alkuosan eri muodot rajaavat yhdyssanan mahdollisia merkitystulkintoja eri tavalla. Suomen kielen yhdyssanoissa avoimimmaksi tulkinnan jättää nominatiivi, joka paljastaa yhdyssanan osien välisestä suhteesta ainoastaan sen, että osat ovat jossakin suhteessa toisiinsa (ks. myös funktionalistisen yhdyssanateorian esittely edellisestä luvusta 2.3.1). Merkitystulkinta on täysin avoin ja perustuu ainoastaan kontekstiin, maailmantietoon ja tietoon kyseisen yhdyssanan vakiintuneesta merkityksestä (esim. Räisänen 1986). Myös yhdysosamuodoista yleisin, *-nen*-loppuisten sanojen *-s*-loppuinen muoto, jättää suhteen laadun avoimeksi (esim. *ihmiskuljettaja* [TT], *hevosteatteri* [TT]).

Nominatiivin jälkeen muista sijamuodoista eniten mahdollisia merkitystulkintoja on genetiivillä. Vesikansan (1989: 224–225; ks. myös jo Genetz 1881: 21–22 sekä Jaakola 2004: 148) mukaan genetiivillä on yhdyssanan alkuosana kaksi mahdollista merkitystehtävää: 1) Possessiivis-subjektiivisessä merkityksessä oleva genetiivimuotoinen alkuosa ilmaisee joko deverbaalin jälkiosan osoittaman toiminnan tekijää (*aamunkoitto*, *lahnankutu*) tai erilaisia possessiivisiä suhteita. Possessiivinen suhde voi liittyä esimerkiksi omistukseen, hallintaan tai yhteenkuuluvuuteen (*kauptungintalo*, *perheenjäsen*), sijaintiin (*korvantausta*, *joenmutka*), alkuperään (*kananmuna*), aiheuttajaan (*auringonpistos*, *bensiininkatku*), tarkoitukseen (*lapsensäntky*, *miestenlehti*) tai seuraukseen (*kuolemansyy*). 2) Objektiivisessä merkityksessä oleva genetiivimuotoinen alkuosa ilmaisee deverbaalin jälkiosan tarkoittaman toiminnan kohdetta (*hinnanalennus*, *aloitteentekijä*, *aallonmurtaja*).

Yhdyssanan alkuosan erilaisten merkitystehtävien tarkka luokittelu on ollut keskeistä erityisesti generativistisessä perinteessä. Yhdyssanojen syvärakenteisiin perustuva jaottelu heijastuu myös edellisessä kappaleessa mainittuun Vesikansan esitykseen erilaisista genetiivimuotoiselle alkuosalle mahdollisista possessiivisistä merkitystehtävistä. Käytännössä tällaisten luokittelujen lopputulos vaihtelee paljon sen mukaan, mistä näkökulmasta tai millaisessa kontekstissa sanaa tarkastellaan. ISK:ssa (§ 411) genetiivimuotoisella substantiivilla alkavan yhdyssanan osien välisen merkityssuhteen mainitaan olevan tapauksittain hyvin erilainen ja suhdetta kuvataan vain hyvin yleisellä tasolla: ISK:n mukaan ”yhdyssubstantiivissa genetiivimuotoinen määriteosa ilmaisee, mihin kokonaisuuteen tai minkä alaan tai piiriin edusosan tarkoite kuuluu”. Genetiivialkuisista yhdyssubstantiiveista ISK (§ 411, 423) erottaa omaksi semanttiseksi tyyppikseen vain sellaiset appositioyhdyssanat, joissa alkuosa ilmaisee yhdyssanan tarkoitteen neutraalisti ja jälkiosa kuvailee sitä (esim. *äijänkäppänä*, *puunkuvatus*; ks. myös Jaakola

2004: 149). Yhdysadjektiivieissa genetiivimuotoinen alkuosa voi ilmaista jälkiosan ilmaisemalle ominaisuudelle tyypillistä vertailukohtaa (*kuvankaunis, langanlaiha*) tai määrittää pakollisen seuralaisen vaativaa suhteutusadjektiivia (*hyvännäköinen, säännönmukainen*; suhteutusadjektiiveista ks. ISK § 270). Lisäksi genetiivimuotoinen adjektiivi voi yhdysadjektiivin määriteosana modifioida jälkiosaa myös jollain muulla tavoin (*helakanpunainen, tyhmänrohkea*) (tyypin merkityksestä ks. myös Orpana 1988: 119–166).

Yhdyssanan genetiivimuotoisen alkuosan merkitystehtävät ovat monelta osin samoja kuin nominatiivin, vaikka osien suhde ei olekaan yhtä avoin ja vapaatulkintainen (ISK § 411). Toisinaan kieleen onkin vakiintunut samaan merkitykseen sekä nominatiivi- että genetiivialkuinen variantti (esim. *salaattikas-tike ~ salaattinkas-tike*). Vain harvoin alkuosan eri muotojen välille on vakiintunut merkitysero (*maamies ≠ maanmies, heikkonäköinen ≠ heikkonnäköinen*). Yhdyssubstantiiveissa vaihtelun taustalla on samojen merkitystehtävien sekä esimerkiksi fonologisten seikkojen lisäksi ilmeisesti myös vanha murre-ero. Länsimurteissa on vanhastaan käytetty genetiivialkua kun taas itämurteissa on suosittu nominatiivialkua. (ISK § 412; ks. kuitenkin myös Penttilä 1963 [1957]: § 223, s. 268; Kapiainen 1983.) Nykyiseen yleiskieleen on yleisemmin vakiintunut nominatiivialkuinen variantti (*jalkapohja ~ jalanpohja*) mutta joskus myös genetiivialkuinen (*järvenranta ~ järviranta*). Vaihtelua esiintyy varsinkin substantiivialkuisissa yhdysubstantiiveissa ja *-inen*-loppuisissa yhdysadjektiivieissa. (ISK § 412–413.)

Muut sijamuodot rajaavat yhdyssanan merkitystä vielä genetiiviä tarkemmin. Niiden rakenne antaa vihjeitä osien keskinäisestä suhteesta tai osoittaa sen samaan tapaan, kuin esimerkiksi sijapäätteet osoittavat sanojen keskinäiset suhteet syntaktisissa rakenteissa. Esimerkiksi yhdyssanat *maahantuoda* (TT) ja *istumaannousu* (TT) ovat yhtä läpinäkyviä kuin vastaavat syntaktiset rakenteet, vaikka ne yhdysanoina ovatkin leksikaalistuneet tarkempaan merkitykseen. Muussa sijamuodossa olevan alkuosan sisältävän yhdyssanan rinnalla esiintyykin yleensä vastaava kaksisanainen lauseke (esim. *kotiinpaluu ~ kotiin paluu* tai *paluu kotiin*, ISK § 418).

Yhdyssanan alkuosan muodon valintaan vaikuttavat toisaalta pyrkimykset rakenteen keveyteen ja toisaalta viestin välittymisen varmistamiseen (vrt. esim. äänne- ja muoto- vaihtelut). Muut sijat kuin nominatiivi ja genetiivi osoittavat yhdyssanan osien välillä olevan merkityssuhteen yksiselitteisesti ja tekevät siten viestin välittymisestä varmempaa. Yleensä nämä muodot kuitenkin pidentävät yhdyssanaa ja voivat näin tehdä rakenteesta raskaan. Koska yhdyssanan nominatiivimuotoinen alkuosa (ja osa yhdysosamuodoista) jättää osien välisen merkitystulkinnan täysin avoimeksi, nominatiivi on sananmuodostuksessa periaatteessa aina mahdollinen vaihtoehto mille tahansa muulle muodolle. Joissain tapauksissa varsinkin yleiskielessä oikeakielisyyden vaatimus voi rajoittaa vapaata valintaa (ks. esim. Vesikansa 1989: 227–237; T. Kolehmainen 2011), mutta sanan merkitystulkintaan liittyviä esteitä ei nominatiivin käytölle ole. Nominatiivialun yleisyys suomen yhdysanoissa viittaa siihen, että konteksti ja maailmantieto yhdessä riittävät ohjaamaan kielenkäyttäjän oikeaan tulkintaan.

Esimerkiksi mainoksissa ja otsikoissa yhdyssanojen potentiaalista monitulkintaisuutta ja sen synnyttämää huumoria käytetään myös tietoisesti (esimerkit 3 ja 4). Räisäsen (1990: 5) mukaan monitulkintaisuuden ja sen synnyttämä merkityksen hämäryyden hyödyntäminen voi olla tietoista myös politiikan kielessä.

- (3) [Roots- eli juurirockyhtye] Gangster of Love kävi juurihoidossa Floridassa [otsikko] [TS 27.3.2014, s. 22]
- (4) Kaupungin parasta yöelämää [sänkyliikkeen tv-mainos] [Liv 24.3.2014, Sinkkuelämää-sarjan mainoskatkolla]

Huomionarvoista on, että alkuosan muodon lisäksi yhdyssanan merkityksen rakentumiseen ja tulkintaan vaikuttaa myös muodosteen osien sanaluokka. Esimerkiksi yhdyssanan alkuosana toimiva adjektiivi osoittaa alkuosan viittaavan jälkiosaan liittyvään ominaisuuteen (esim. *yhteisauto* [TT], *kesyhilleri* [TT]). Verbimuoto puolestaan kertoo, että alkuosa viittaa jälkiosaan liittyvään tekemiseen (esim. *nukkumaparvi* [TT], *raastinpuikko* [TT]). Substantiivit jättävät tulkinnan avoimemmaksi, sillä ne voivat viitata sekä ominaisuuteen, tekemiseen että johonkin muuhun jälkiosaan liittyvään asiaan (esim. tekemiseen viittaavat alkuosat sanoissa *avautumissivu* [TT], *kuvailutulkkaus* [TT]). Substantiivit voivat siis yhdyssanan alkuosana toimia paitsi substantiivisissa myös adjektiivisissa ja verbaalisissa merkitystehtävissä.

2.3.3 Kirjoitettujen yhdyssanojen prosessointi

Tässä luvussa käsittelen yhdyssanojen prosessointia ja erityisesti sitä, miten lukija päätyy muodostamaan käsityksen kirjoitetun yhdyssanan merkityksestä. Olennainen ero kirjoitetun ja puhutun kielen prosessoinnin välillä on se, että tekstiä lukiessaan ihmisellä on mahdollisuus palata taaksepäin. Tärkeimpiä yhdyssanojen prosessointiin ja tulkintaan liittyvässä tutkimuksessa esitettyjä kysymyksiä on, tulkitaanko morfologisesti kompleksiset sanat (yhdyssanat ja johdokset) kokonaisuuksina vai osa kerrallaan. Kysymykseen on esitetty käytännössä kolmenlaisia vastauksia.

Yhdyssanojen kiteytyneisyyden kannattajista pisimmälle menevät niin sanotun holistisen näkemyksen kannattajat, joiden mukaan yhdyssanat säilötään muistiin ilman mitään tietoa niiden morfologisesta rakenteesta. Heidän mukaansa yhdyssanat siis myös tuotetaan ja tulkitaan kokonaisuuksina. Holistisen näkemyksen kannattajat eivät katso tässä tutkimuksessa keskeisellä yhdyssanojen sisäisellä morfosyntaktisella rakenteella tai sen enempää semanttisellakaan rakenteella olevan vaikutusta muodosteiden tulkintaan. (Tarkemmin näitä näkemyksiä esittelevät esim. Hyönä–Pollatsek 1998; Pollatsek–Hyönä–Bertram 2000.)

Monien psykologingvistien suorittamien kokeiden perusteella vaikuttaa siltä, että teoreettisesti rakenneltu holistinen näkemys yhdyssanojen olemuksesta on psykologisesti vajavainen (tosin ks. esim. van Jaarsveld–Rattink 1988; de Jong ym. 2002). Toisen, niin sanotun atomistisen näkemyksen kannattajat esittävätkin, että yhdyssanat on jaettava semanttisesti läpinäkyviin ja opaakkeihin. Näistä läpinäkyvät yhdyssanat prosessoidaan osina. Näkemystä tukevat monet psykologingvistiset

reaktioaika- ja silmänliiketutkimukset. (Esim. Marslen-Wilson–Tyler–Waksler–Older 1994; Fiorentino–Poeppel 2007.) Tutkimuksissa on saatu jonkinasteista näyttöä myös siitä, että jopa semanttisesti opaakkien yhdyssanojen prosessointiin kuuluu sanojen jakaminen osiin ortografisella tai vähintään semantiikasta erillisellä morfologisella tasolla (Jarema ym. 1999; Pollatsek–Hyönä 2005). Käsitystä tukevat myös afasiapotilaiden tutkimuksesta saadut tulokset (Ulfsbjorninn 2009). Nähdäkseni tätä käsitystä tukee myös se, että suomen kielessä myös opaakkien yhdyssanojen osien oikeinkirjoitussääntöjen vastainen erilleen kirjoittaminen on yleistä³¹.

Kolmannen, atomistisen ja holistisen näkemyksen väliin jäävän kannan mukaan morfologisesti kompleksisten sanojen osina ja kokonaisuutena tulkittaminen kilpailevat keskenään. Tämän niin sanotun *race model* -mallin mukaan morfologisesti kompleksisen sanan tulkinnessa käytetään yhtäaikaaisesti molempia reittejä. Lopullisesti tulkinta tehdään yleensä sen reitin kautta, joka johtaa nopeammin lopputulokseen. (Hyönä–Pollatsek 1998; Pollatsek–Hyönä–Bertram 2000.) Tulkittamisenopeuteen vaikuttavat monet seikat. Koko muodosteen pituuden lisäksi olennaisia ovat esimerkiksi koko muodosteen ja sen osien frekvenssi sekä osien pituus (Laudanna–Burani 1995; Bertram – M. Laine – K. Karvinen 1999; Bertram–Hyönä 2003; Hyönä–Bertram–Pollatsek 2004). Yhdyssanan tulkittamisenopeuteen ja sitä kautta myös tulkittamistapaan vaikuttavista tekijöistä tässä tutkimuksessa tarkastellaan lähemmin suomen kielen uusien yhdyssanojen ja niiden osien pituutta (luku 5.1).

Tekstiä lukevan ihmisen toiminnallisen havaintokentän (*perceptual span*) on arvioitu ulottuvan (vasemmalta oikealle luettavissa kielissä) maksimissaan noin 15 merkkiä oikealle ja 4 merkkiä vasemmalle katseen kiinto- eli fiksaatiopisteestä. Sanojen merkityksiä ihminen pystyy kuitenkin tulkitsemaan vain noin 12 merkin laajuiselta alueelta (*word identification span*), joka ulottuu oikealta vasemmalle luettavissa kielissä noin 8 merkkiä fiksaatiopisteestä oikealle ja 4 merkkiä vasemmalle. Tarkan näön eli fovean alueelta yli jäävältä toiminnallisen näkökentän alueelta (alue 8–15 merkkiin oikealle) lukija poimii tietoa sanojen pituuksista ja visuaalisesta muodosta. Näiden tietojen avulla lukija pystyy ohjaamaan katseensa seuraavan fiksaatiopisteen mahdollisimman keskelle seuraavaa vielä tunnistamatonta sanaa, jolloin mahdollisimman suuri osa uudesta sanasta osuu tarkan näön alueelle. (Rayner–Pollatsek 1989; Hyönä 1995: 104.)

Sanojen pituudesta riippuen tarkan näön alue kattaa yleensä 1–2 sanaa. Käytännössä tekstiä ei kuitenkaan aina lueta näin suurin harppauksin. Suomea lukevan aikuisen lukijan keskimääräinen silmänliikkeen eli sakkadin pituus on noin 11 kirjainta (Hyönä – P. Niemi 1990; englantia luettaessa sakkadin pituus on keskimäärin 7–9 merkkiä, ks. Rayner 1998: 375). Esimerkiksi 12 merkkiä pitkät yhdyssanat luetaankin suomessa tyypillisesti kahdella fiksaatiolla (Hyönä–Pollatsek 1998; Hyönä–Bertram–Pollatsek 2004; ks. myös Hyönä – P. Niemi – Underwood 1989). Sen sijaan

31 Välttämättä kyse ei tarvitse olla edes varsinaisesta yhdyssanasta, vaan ilmauksen hahmottuminen kaksiosaiseksi saattaa riittää. Esim. *mikä olisi hyvä ekaksi auto maatti aseeksi? pitäisi olla alle 150e* (Suomi24; Mikä ekaks; ks. myös Niinimäki 1991: 24–25).

8 merkkiä pitkät yhdyssanat voidaan lukea myös yhdellä fiksaatiolla (Bertram–Hyönä 2003: 624–625).

Mikäli yhdyssana on kokonaisuutena niin pitkä, että sen lukeminen vaatii vähintään kaksi fiksaatiota, todennäköisyys muodosteen tulkitsemiseen osina lisääntyy (Bertram–Hyönä 2003; Hyönä 2012). Tällaisissa sanoissa osien tunnistamista helpottavat vihjeet (esim. yhdysviivat) nopeuttavat sanan prosessointia. Sen sijaan lyhyet, kerralla tarkan näön alueelle mahtuvat sanat tulkitaan todennäköisemmin kokonaisuuksina. Niissä osien jakoa osoittavat vihjeet hidastavat prosessointia, sillä ne rohkaisevat lukijaa valitsemaan kokonaisuuden kautta tulkitsemisen sijaan hitaamman osien kautta kulkevan reitin. (Hyönä 2012; Bertram–Hyönä 2013; ks. myös Bertram–Pollatsek–Hyönä 2004.) Pitkissä yhdyssanoissa myös yhteen kirjoitetun osuuden pituus vaikuttaa ilmauksen prosessointiin. Inhoff, Radach ja Heller (2000) ovat tutkimuksessaan havainneet, että pitkän yhdyssanan osien erilleen kirjoittaminen nopeuttaa yhdyssanan osien tunnistamista ja siten nopeuttaa yhdysanojen lukemista (jopa kielissä, joissa se on oikeinkirjoitussääntöjen vastaista; Inhoffin, Radachin ja Hellerin tutkimus koskee saksaa). Samalla osien erikseen kirjoittaminen kuitenkin hankaloittaa käsitteellisesti yhtenäisen yhdyssanan merkityksen hahmottamista.

Suomessa kielenhuollon suositus on pitkään ollut, että tulkinnanvaraisissa tapauksissa yhdyssanan osat on parempi kirjoittaa erikseen kuin yhteen. Ohje koskee muilla sijamuodoilla kuin nominatiivilla alkavia yhdyssanoja. Ohjetta on perusteltu pitkien yhdyssanojen hankalalla hahmottuvuudella ja epäesteettisyydellä. Suositus saa tukea myös edellä esitellyistä yhdyssanojen prosessointia koskevista tutkimuksista, sillä pitkien (yhdyssanojen) tulkitseminen on hitaampaa ja vaikeampaa kuin lyhyiden. Merkkimäärällä mitatun pituuden lisäksi tulkintaan vaikuttaa myös se, kuinka kompleksisesta muodosteesta on kyse (johdosten osalta Brattico–Leinonen–Krause 2007).

Yhdyssanan alkuosan muodon vaikutusta yhdyssanan prosessointiin ei ole tois-
taiseksi tutkittu (osien välisten merkityssuhteiden ymmärtämistä ovat tutkineet mm. Gagné–Shoben 1997; Gagné 2001; jälkiosan kantasanan sanaluokan vaikutusta on käsitellyt esim. Mäkisalo 2000a: 171–179). Alkuosan muoto tarjoaisi sikäli mielenkiintoisen psykolingvistisen tutkimuskohteen, että toisaalta muut sijat kuin nominatiivi ja genetiivi tekevät yhdyssanoista yleensä pidempiä mutta toisaalta ne myös osoittavat nominatiivia ja genetiiviä yksiselitteisemmin osien välisen merkityssuhteen (ks. myös edellinen luku 2.3.2).

3 Produktiivisuus

Tässä luvussa käsittelen toista tutkimukseni kahdesta tärkeimmästä käsitteestä – produktiivisuutta. Ensimmäisessä alaluvussa (3.1) tarkastelen produktiivisuuden käsitettä. Luvussa esittelen produktiivisuudesta aiemmassa tutkimuksessa esitetyjä näkemyksiä sekä kuvaan tässä tutkimuksessa käytettävän produktiivisuuden määritelmän. Toisessa alaluvussa (3.2) esittelen tekijöitä, joiden on aiemmassa tutkimuksessa havaittu vaikuttavan produktiivisuuteen. Tarkastelen aihetta erityisesti yhdyssanojen näkökulmasta. Kolmannessa alaluvussa (3.3) kuvaan produktiivisuuden mittaamisessa aiemmin käytettyjä metodeja sekä esittelen tässä tutkimuksessa käytetyn uuden metodin.

3.1 Produktiivisuus käsitteenä

Produktiivisuus on keskeinen sananmuodostukseen liittyvä käsite, jolla tarkoitetaan kielellisen rakenteen kykyä tuottaa uusia kielenmukaisia ilmauksia. Tästä perusmääritelmästä lienevät lähes kaikki tutkijat samaa mieltä, mutta määritelmän tarkemmista yksityiskohdista ei ole kielitieteessä yksimielisyyttä. Produktiivisuutta käsittelevässä tutkimuksessa onkin otettava kantaa ainakin neljään tärkeään kysymykseen, joita käsittelen tämän luvun alaluvuissa (ks. myös Bauer 2001: 32): 1) mikä kielessä on produktiivista (3.1.1), 2) onko tarkastelun kohteena rakenteen kieliopillinen potentiaali uusien muodosteiden tuottamiseen vai sen tosiasiallinen suosio kielenkäytössä, 3) tarkastellaanko produktiivisuutta diakronisesta vai synkronisesta näkökulmasta (kysymykset 2 ja 3 luvussa 3.1.2) ja 4) nähdäänkö produktiivisuus totaalisisena vai asteittaisena ilmiönä ja millaisia asteita siitä erotetaan (3.1.3). Näiden kysymysten voidaan nähdä liittyvän suoraan siihen, miten produktiivisuuden käsite ymmärretään. Esimerkiksi Laurie Bauerin (2001: 1, 32) mukaan erilaiset näkökulmat kuitenkin ennemmin täydentävät toisiaan kuin ovat keskenään ristiriitaisia.

3.1.1 Mikä on produktiivista?

Produktiivisuuden käsitteestä on sekä fennistiikassa että kansainvälisessä produktiivisuuden tutkimuksessa puhuttu yleisimmin johto-opin yhteydessä. Produktiivisuutta on käytetty myös johtamisen ja taivutuksen erottamiskriteerinä. Taivutus on yleensä täysin produktiivista eli mahdollista kaikista tietyn sanaluokan sanoista. Johtaminen sen sijaan on harvoin täysin produktiivista. (Esim. ISK § 61.)

Produktiivisuuden käsite sopii myös muiden kielen tasojen kuin sananmuodostuksen kuvaamiseen (esim. syntaksiin liittämisestä ks. esim. Chomsky 1965: 5), ja laajimmillaan produktiivisuus onkin nähty koko kielisysteemin ominaisuutena (Bauer 2001: 12). Produktiiviseksi eli uutta luovaksi on kielisysteemissä voitu käsitellä niin affiksit, (morfologiset) prosessit tai useista prosesseista koostuvat joukot (esim. Anderson 1982: 585; Bauer 2001: 32) kuin säännötkin (esim. Aronoff 1976: 36).

Kun tietty morfologinen prosessi toistuu tarpeeksi monta kertaa, sitä voidaan kuvata myös sääntönä. Monien tutkijoiden mukaan produktiivisia voivatkin olla vain säännöillä kuvattavissa olevat ilmiöt³² (esim. Rintala 1978: 162; Bauer 2001: 5). Käytännön tutkimustyön kannalta sekä prosessien että sääntöjen ongelma on kuitenkin niiden abstraktius. Prosesseihin ja sääntöihin voidaan päästä käsiksi oikeastaan ainoastaan kielen käytössä esiintyviä rakenteita tutkimalla. Rakenteen taajuus sanastossa toki heijastaa myös kyseistä rakennetta noudattavia sanoja tuottavan prosessin produktiivisuutta, mutta tämän työn kaltaisessa valmiita sanoja tarkastelevassa tutkimuksessa on silti luontevampaa käsitellä produktiivisuutta suoraan tutkimusaineistossa näkyvien rakenteiden kuin jo aiemmin tapahtuneiden prosessien ominaisuutena.

Johto-opin kuvauksissa produktiivisuus on varsinkin aiemmin usein käsitetty johtimen tai yleisemmin affiksin ominaisuudeksi (esim. Penttilä 1963 [1957]: § 228; Rintala 1978: 153; ks. myös Bauer 2001: 12), mutta nykyään produktiivisuus esitetään yleensä ennemmin johdostyyppin kuin johtimen ominaisuutena (ks. esim. VISK määritelmät s. v. *produktiivisuus*; ISK § 164). Johdostyyppi käsittää vakio-osan eli johtimen lisäksi avoimen paikan tietyt ehdot täyttävälle leksikaaliselle elementille (johdoksen kantasanalle, esim. [verbivartalo + *-minen*]). Yleensä ehdot koskevat kantasanan sanaluokkaa, mutta johdostyyppi voi edellyttää myös kantasanan kuumista johonkin tiettyyn semanttiseen ryhmään. (ISK § 148, 164; ISK:ssa johdostyyppjä kuvataan myös muottien avulla, § 148).

Mikäli johdosten tutkimuksesta tuttu ajatus produktiivisuudesta johtimen ominaisuutena siirrettäisiin yhdyssanojen tutkimukseen, se vastaisi produktiivisuuden käsitteistä yhdyssanan osana toimivien sanojen ominaisuudeksi.

32 Monet tutkijat pitävät osoituksena epäproduktiivisuudesta myös sitä, että rakenteen tuottamat uudet sanat ovat tietoisesti muodostettuja (tarkemmin esim. Bolozky 1999: 1–2). Rajaa tietoisien ja tiedostamattoman sananmuodostuksen välille on kuitenkin vaikea vetää esimerkiksi siksi, että eri kielenkäyttäjät tiedostavat kielen elementit eri tavoin (Plag 1999: 13–14; Bauer 2005: 317–318). Tässä tutkimuksessa en tee eroa tietoisesti ja tiedostamatta muodostettujen uudissanonjen välille.

Johdostutkimuksessa sillä, pidetäänkö produktiivisuutta johtimen vai johdostyyppin ominaisuutena, ei ole suurta käytännön merkitystä. Yhdyssanatutkimuksessa rakenteiden ja yhdyssanan osina toimivien sanojen produktiivisuuden erottaminen on sen sijaan välttämätöntä: Yhdyssanojen muodostaminen on sekä kieliopillisesti että semanttisesti hyvin vapaata ja yksittäisten sanojen valintaan yhdyssanan osaksi vaikuttaa todennäköisesti eniten se, millaisista asioista kielenkäyttäjillä on tarve puhua. Näin ollen käsitys produktiivisuudesta pelkää yksittäisten sanojen ominaisuutena ohjaisi tutkimuksen kielen sijasta kielenulkoiseen maailmaan. Yhdyssanatutkimuksessa produktiivisuutta on siis pidettävä enemmän yhdyssanarakenteen kuin yksittäisten sanojen ominaisuutena. (Ks. myös Mäkisalo 2000a: 13–14.)

Tässä tutkimuksessa käsittelen produktiivisuutta ensisijaisesti morfosyntaktisen yhdyssanarakenteen kyknä tuottaa uutta sanastoa (ks. myös Mäkisalo 2000a: 13–14). Morfosyntaktisella yhdyssanarakenteella tarkoitan abstraktia rakennetta, jossa on kaksi tai useampia vapaita paikkoja tietyt ehdot täyttävillä leksikaalisille elementeille. Vapaille paikoille sopivia elementtejä koskevat ehdot liittyvät lähinnä sanaluokkaan ja muotoon, mahdollisesti myös pituuteen. Esimerkiksi rakenne [SUBST.GEN.]+[ADJ.] tuottaa substantiivista ja adjektiivista koostuvia kaksi leksikaalista osaa sisältäviä yhdyssanoja, joiden alkuosa on genetiivissä (esim. *arjenkarhea* [TT]; eri rakenteiden vaikutuksesta yhdyssanan merkitykseen ks. luku 2.3.2).

Laajan analogiakäsityksen³³ mukaan kaikki sananmuodostus pohjaa analogiaan eli pyrkimykseen luoda sanoja, jotka ovat rakenteeltaan olemassa olevien sanojen kaltaisia (sananmuodostuksen ohella analogia vaikuttaa myös kaikkeen muuhun kielenkäyttöön; analogiasta ja sitä koskevista teorioista tarkemmin esim. Esa Itkonen 2005). Olemassa olevat rakenteet ovat siis sananmuodostuksessa morfologisten prosessien lähtökohta. Analogioihin perustuva sananmuodostuskäsitys ei tarkoita sitä, että produktiivisina pidettäisiin kaikkia rakenteita, jotka tuottavat yhdenkin uuden muodosteen (ks. myös ISK § 164). Ollakseen produktiivinen rakenteen on tuotettava uutta sanastoa toistuvasti (produktiivisuuden asteista ks. tarkemmin luku 3.1.3).

Morfosyntaktisen yhdyssanarakenteen produktiivisuuden kriteerinä on lisäksi pidettävä sitä, että rakennetta noudattavissa uusissa sanoissa esiintyy useita erilaisia leksikaalisia elementtejä. (Ks. tämän tutkimuksen luvut 5–7). Mikäli esimerkiksi kaikkien suomen kieleen syntyvien uusien yhdysadjektiivien jälkiosa olisi sana *-ystävällinen* (täysin kuvitteellinen esimerkki), produktiiviseksi olisi katsottava yhdysadjektiivien yleisen morfosyntaktisen rakennetyypin sijaan vain *-ystävällinen*-loppuiset yhdysadjektiivit, jolloin produktiivisuus voitaisiin ymmärtää myös pelkän *-ystävällinen*-loppuosan (eli yksittäisen sanan) produktiivisuudeksi (vrt. produktiivisuuden ymmärtäminen johtimen ominaisuudeksi). Rakenteellisella tasolla epäproduktiiviseksi on katsottava myös sellaiset rakenteet, joissa kaikkien rakennetta noudattavien uusien yhdyssanojen taustalta löytyy jonkinlainen

33 Suppean analogiakäsityksen mukaan analogialla on ymmärretty ainoastaan säännöllisistä rakenteista poikkeamista (ks. esim. A. Hakulinen – Ojanen 1976 s.v. *analogia*).

poikkeuksellinen muodostusmotivaatio (esim. käännöslainaus tai vanhan sanan käyttö uudessa merkityksessä). Koska morfosyntaktisten yhdyssanarakenteiden produktiivisuuden havaitseminen ja varmistaminen vaatii myös sanakohtaisen produktiivisuuden tarkastelua, tässä tutkimuksessa kvantitatiivista analyysia tukee läpi työn kvalitatiivinen tarkastelu. Lisäksi eri yhdyssanarakenteiden leksikaalista monipuolisuutta on tarkasteltu tilastollisesti luvussa 6.2.

3.1.2 Kieliopillinen ja käytännöllinen sekä diakroninen ja synkroninen produktiivisuus

Tässä tutkimuksessa erotan produktiivisuudesta kaksi puolta: kieliopillisen potentiaalin eli sen, mikä on mahdollista, ja käytännön suosion eli sen, mikä toteutuu käytännössä. Kieliopilliselta kannalta katsottuna olennaista on se, mikä on kielen kannalta ylipäänsä mahdollista eli mitkä ovat rakenteen kieliopilliset mahdollisuudet uusien kielenmukaisten muodosteiden tuottamiseen. Käytännön suosion kannalta olennaista taas on se, kuinka paljon rakenteen avulla tuotetaan uusia muodosteita todellisessa kielenkäytössä. Tässä tutkimuksessa huomion kohteena ovat rakenteen todelliset uudet toteumat eli käytännöllinen produktiivisuus. Viittaan tässä tutkimuksessa rakenteen käytännölliseen produktiivisuuteen myös termeillä *rakenteen suosio* ja *suosittuus* (esimerkiksi Kasik [1980: 60–63] käyttää termiä *aktiivisuus*).

Monissa produktiivisuuden määritelmässä korostetaan kieliopillista produktiivisuutta, josta rakenteen käytännön suosio nähdään vain välillisenä osoituksena. Esimerkiksi Bauerin (2001: 21) mukaan rakenteen tuottamat uudet muodosteet ovat ennemminkin produktiivisuuden indikaattori kuin produktiivisuutta itseään. Koska varsinainen ilmiö voidaan kuitenkin tunnistaa vain sen indikaattorin kautta, pidän teoreettisesti kestävimpanä ratkaisuna kieliopillisen ja käytännöllisen produktiivisuuden erottamista toisistaan. Tämän työn kaltainen empiirinen tutkimus ottaa kantaa vain käytännölliseen produktiivisuuteen.

Rakenteen kieliopillinen potentiaali uuden tuottamiseen ja sen käytännön suosio eivät välttämättä kulje käsi kädessä. Käytännön suosiota ei voi syntyä ilman kieliopillista potentiaalia, mutta joskus rakenne, jolla olisi kaikki kieliopilliset edellytykset tuottaa uutta, ei vain saa suosiota kielenkäyttäjien keskuudessa (Bauer 2001: 21; ks. myös Plag 1999: 6–11; rakenteen suosioon vaikuttavista tekijöistä ks. luku 3.2). Kuten olen jo edellä huomauttanut, kaikki rakennetta noudattavat uudet muodosteet eivät myöskään ole osoitus juuri tutkittavan rakenteen käytännöllisestä produktiivisuudesta: produktiivisuutta arvioitaessa on kiinnitettävä tarkkaa huomiota siihen, mihin suosio itse asiassa kohdistuu (esim. edellisen 3.1.1-luvun kuvitteellinen esimerkki *-ystävällinen*-loppuisista yhdysadjektiveista ja yhdysadjektiveista ylipäänsä).

Produktiivisuutta on perinteisesti tarkasteltu pääasiassa kieliopin näkökulmasta (Baayen 1992: 109), vaikka useimmissa varhaisissakin tutkimuksissa on otettu huomioon vähintään joitain kielenkäyttöön liittyviä pragmasemanttisia rajoituksia (ks. esim. Rintala 1978: 154). Tällaisessa tarkastelussa voi ajatella ainakin synkronisessa näkökulmassa metodiksi riittävän pelkän tutkijan oman kielitajun. Kuten

muussakin kielitieteessä, myös produktiivisuuden tutkimuksessa on viime aikoina kuitenkin herännyt kiinnostus myös todellisen kielenkäytön paremmin huomiointiin näkökulmiin ja metodeihin. Erityisesti uuden näkökulman ja metodien kehittämisessä on kunnostautunut Harald Baayen (tarkemmin luku 3.3 ja esim. Baayen 1992, 1993; Baayen–Lieber 1991; Baayen–Renouf 1996; ks. myös Mäkisalo 2000a: 13). Käytännöllisen produktiivisuuden tarkastelussa pelkkä intuitio ei riitä, vaan mukaan on otettava aineisto, jota on tarkasteltava tilastollisesti.

Produktiivisuus ei ole kielen rakenteiden pysyvä ominaisuus, vaan rakenteiden suosio ja kyky tuottaa uutta vaihtelevat ajan kuluessa: se mikä oli produktiivista joskus aiemmin, ei välttämättä ole produktiivista enää nykyään (Rintala 1978: 155; Baayen 1992: 110–111; Bauer 2001: 6–9). Kielen muutoksen kannalta rakenteiden tämänhetkinen suosio on kieliopillista potentiaalia olennaisempaa, sillä tiheästi kielessä esiintyvistä rakenteista otetaan todennäköisesti mallia myös uusiin muodosteisiin, kun taas kieliopin kannalta yhtä mahdollinen mutta käytännössä harvinaisen rakenteen on suuremmassa vaarassa kadottaa produktiivisuutensa kokonaan. Tämän tutkimuksen näkökulma on synkroninen, eli tarkoituksena on selvittää, mitkä rakenteet ovat tuottaneet uusia muodosteita juuri tarkastelun kohteena olevana aikana.

Diakronisesta näkökulmasta tarkasteltuna produktiivinen voi olla myös sellainen rakenne, jonka kyky uuden luomiseen on jo päättynyt. Osa vanhoista rakenteista voi kivettyä tai kadota kokonaan, ja käyttöön tulee myös kokonaan uudenlaisia rakenteita. Joskus myös jo produktiivisuutensa menettänyt rakenne voi saada sen takaisin (ks. esimerkkejä pääasiassa englannista mutta myös muista Euroopan indoeurooppalaisista kielistä esim. Bauer 2001: 6–10). Produktiivisuuden palautumiseen voi olla useita syitä. Yksi tie rakenteen uuteen produktiivisuuteen voi olla sen elvyttäminen tietoisessa sanastotyössä (L. Hakulinen 1979: 472–473). Suomessa produktiivisuutensa jo menettäneitä rakenteita käytettiin tietoisessa sanastotyössä erityisesti 1800-luvulla, jolloin uuden sanaston tarve oli erityisen suuri. Nykyään sen sijaan tietoisessakin sanastotyössä pyritään pitäytymään muun kielen asettamissa rajoissa. (Rintala 1978: 155.)

Fennistisessä johto-opin kuvauksessa (johon produktiivisuuden käsite siis suomalaisessa tutkimuksessa ensisijaisesti liittyy) ei ole aina selvästi erotettu yhä jatkuvaa ja jo päättynyttä produktiivisuutta, vaan kieltä muuten synkronisesta näkökulmasta tarkastelevissakin esityksissä on saatettu luetella kieleen kuuluvat johtimet diakronisesta näkökulmasta (ks. tarkemmin esim. Rintala 1978: 151). Oma tutkimukseni ei millään tavalla kiistä diakronista näkökulmaa, mutta synkronisen näkökulmani vuoksi yhä jatkuvan ja jo päättyneen produktiivisuuden erottaminen toisistaan on omassa tutkimuksessani tärkeää. Kun ero on tehty, synkronisen kuvauksen kautta saatuja tietoja voidaan myöhemmin käyttää myös diakronisen kuvauksen apuna. Esimerkiksi Bauer (2001: 9) korostaa, että produktiivisuuden kuvaus on tärkeä osa kielen morfologian synkronista kokonaiskuvaa ja muutokset rakenteiden produktiivisuudessa tulisi ehdottomasti huomioida myös kielten diakronisissa kuvauksissa.

3.1.3 Produktiivisuus asteilmiönä

Yksi produktiivisuuden määrittelyn ongelmista on, missä kulkee raja produktiivisuuden ja epäproduktiivisuuden välillä. Analogian ansiosta periaatteessa mikä tahansa muodoste voi toimia mallina uudelle rakenteeltaan samankaltaiselle muodosteelle, mutta ovatko kaikki rakenteet näin ollen ainakin jonkin verran produktiivisia? Yleensä asiaa ei ole käsitetty näin, vaan määritelmässä on pyritty erottamaan jonkinlaisen säännön tai muotinmukaisuuden ohjaama produktiivisuus sellaisesta uusien muodosteiden synnystä, jossa luodaan yksittäisiä uusia muodosteita pelkän (yhteen aiempaan sanaan perustuvan) analogian kautta.

Tutkimuksessa yleisesti käytetty määritelmä produktiivisuuden ja epäproduktiivisuuden rajasta on, että produktiivisen rakenteen kyky synnyttää uusia muodosteita voidaan kuvata jollakin morfologisella ja/tai syntaktis-semanttisella säännöllä kun taas epäproduktiivinen rakenteen kuvaus vaatii kaikkien rakennetta noudattavien tai siinä potentiaalisten tapausten luettelemista erikseen (esim. Rintala 1978: 162; Bauer 2001: 5). Periaatteellinen eronteko produktiivisten ja epäproduktiivisten rakenteiden välille on tällä keinolla mahdollista, mutta käytännössä raja kuitenkin on harvoin selvä (Rintala 1978: 153; Bauer 2001: 4), sillä käytännössä rakenteen produktiivisuuden rajoja kuvaavasta säännöstä voidaan tehdä kuinka yksityiskohmainen tahansa.

Produktiivisten ja epäproduktiivisten rakenteiden erottamisen lisäksi ongelmallista on produktiivisuuden mittaaminen, sillä uutta tuottavistakin rakenteista toiset ovat selvästi produktiivisempia kuin toiset (Rintala 1978; Bauer 2001: 4). Monet tutkijat ovat erottaneet produktiivisuudesta esimerkiksi sellaisia asteita kuin täysin produktiivinen, jossain määrin produktiivinen, marginaalisesti produktiivinen ja epäproduktiivinen. Eri tutkijat ovat erottaneet produktiivisuudesta eri määrän asteita (fennistiikassa esim. Penttilä 1963 [1957]: § 228; Erkki Itkonen 1966: 243; ks. myös Bauer 2001: 125–126), ja ainakin Bauerin (2001: 32, 125–126) mukaan tutkijoiden kesken esiintyy aiheesta aitoa erimielisyyttä. Toisten mielestä kaksi tai kolme astetta riittää, kun taas toiset erottavat produktiivisuuden asteita äärettömän määrän. Produktiivisuuden asteiden äärettömän määrän voi käsittää myös aitona jatkumona, jossa eri asteiden välillä ei ole tarkkoja rajoja (ks. myös luku 3.3.2).

Tässä tutkimuksessa erotan toisistaan produktiiviset ja epäproduktiiviset yhdyssanarakenteet rakenteen tyyppifrekvenssin sekä luvussa 3.1.1 kuvattujen periaatteiden mukaisesti. Varsinkin uudessa sanastossa harvinaisten rakenteiden kohdalla tarkastelen lähemmin myös yksittäisten yhdyssanojen muodostusmotivaatiota ja leksikaalista monimuotoisuutta. Lisäksi kuvaan yhdyssanarakenteiden produktiivisuutta jatkumoluonteisesti vertailemalla rakenteita keskenään. Vertailussa käyttämistäni mittareista tärkein on rakenteen suhteellinen tyyppifrekvenssi uudessa yhdyssanastossa (ks. tarkemmin luku 3.3.4).

3.2 Produktiivisuuteen vaikuttavia tekijöitä

Tässä luvussa käsittelen sekä ylipäänsä yhdyssanojen muodostamisen produktiivisuuteen että tiettyjen yhdyssanarakenteiden produktiivisuuteen vaikuttavia tekijöitä. Kielellisen rakenteen produktiivisuutta ja sen syitä tutkittaessa huomio on syytä kiinnittää ensimmäiseksi rakenteen itsensä ominaisuuksiin (Plag 2003: 61). Jotta kielen rakenne voisi olla produktiivinen, sen on ensinnäkin oltava mahdollinen. Varsinkin produktiivisuutta kieliopillisena potentiaalina tarkastelevissa tutkimuksissa aihetta on usein kuvattu rajoittamisen näkökulmasta (alaluku 3.2.1, ks. myös esim. Rintala 1978; Plag 1999: 37–61, 2006: 550–551).

Kielioppi ja semantiikka eivät yksinään riitä selittämään rakenteiden suosiota. Aina loogista selitystä ei tarjota myöskään rakenteen aiempi suosio. (Bauer 2001: 5–10.) Lisäksi rakenteen käytännölliseen produktiivisuuteen vaikuttavat myös kognitiiviseen prosessointiin liittyvät seikat (luku 3.2.2) sekä sosiaaliset ja tyylilliset seikat (luku 3.2.3).

3.2.1 Kieliopilliset, semanttiset ja leksikaaliset rajoitukset

Sananmuodostuksessa kielen morfologia ja kielioppi asettavat teoriassa reunaehdot sille, mikä on kielessä mahdollista. Vaikka käytännössä rajoja voidaan rikkoa, epäkieliopilliset rakenteet eivät voi olla kovin produktiivisia. Ajan myötä alkuaan epäkieliopillinenkin rakenne voi muuttua produktiiviseksi, mutta sitä ennen se muuttuu väistämättä myös osaksi kieliopillista normia. Koska kielioppi muuttuu hitaasti, synkronista produktiivisuutta tarkastelevassa tutkimuksessa kieliopin asettamia reunaehtoja ei ole järkevää kyseenalaistaa tai olettaa norminvastaisten rakenteiden produktiivisuutta.

Suomen kielioppi ja morfologia asettavat yhdyssanojen muodostukselle vain vähän rajoituksia (ks. luku 2.2.3). Morfologis-kieliopillisten rajoitusten lisäksi sananmuodostusta rajoittavat myös semanttiset ja leksikaaliset seikat. Sananmuodostuksen semanttiset rajoitukset eivät oikeastaan ole kielessä vaan kielenulkoisessa maailmassa ja logiikassa. Suomen kielessä periaatteessa mitkä tahansa sellaiset sanat, joilla on kielenulkoisessa maailmassa jotain tekemistä toistensa kanssa, voidaan yhdistää yhdyssanaksi. Tietyntylaisia semanttisia suhteita sisältävät yhdyssanat kuitenkin vakiintuvat kielenkäyttöön vain harvoin (ks. tarkemmin Downing 1977). Esimerkiksi loogisia vastakohtasuhteita sisältävät sanat ovat harvinaisia, mutta niitä voidaan silti käyttää kielessä kertamuodosteina, sillä viime kädessä vasta konteksti ratkaisee sanojen kulloisenkin merkityksen. Vaikkapa ensi näkemältä loogisesti mahdottomalta vaikuttavan vastakohtaisuuden sisältävä *sinkkuvaimo* voi kielenkäytössä viitata naimisissa olevaan mutta naimattoman tavoin käyttäytyvään tai naimattomaan mutta vaimon tavoin käyttäytyvään naiseen³⁴ (ks. myös Tyysteri

34 Esim. ”Jussi Parviaisen *sinkkuvaimo* väittää saaneensa tekstiviestejä kolmannelta osapuolelta, joka väittää, että takavuosien tunnettu kaunotar Anne Hedman olisi pyytänyt Saanaa toimimaan dominana” (Metropoli: A. H. pyysi S/M-seksiä) ja ”Sunnuntai lapsiperheessä on lapsettomalle *sinkkuvaimolle* harvinaista herkkua” (Sounio 2010).

2011a). Yhdyssanojen muodostukseen liittyvät semanttiset rajoitukset vaikuttavat lähinnä yksittäisten sanojen vakiintumismahdollisuuksiin. Sen sijaan kokonaisen rakenteen produktiivisuuteen niillä ei ole vaikutusta.

Yhdyssanojen muodostusta rajoittavat kahdenlaiset leksikaaliset seikat. Ensinnäkin vaikutusta on kieleen jo vakiintuneilla yhdyssanoilla, sillä kielellä on taipumus välttää toisaalta homonyymejä ja toisaalta jossain määrin myös sellaisia uudissanoja, joiden tarkoitteilla on jo vakiintunut nimitys (Aronoff 1976: 43; van Marle 1986: 607, 613–617; Plag 1999: 50). Esimerkiksi sanaa *naissankari* käytetään kyllä toisinaan viittaamassa naiseen, joka on sankari. Koska sana kuitenkin on vakiintunut ensisijaisesti merkitykseen 'naistenmies, casanova' (KS s. v. *naissankari*), *nais-sankarin* todennäköisyys vakiintua myös naista tarkoittavaksi on pienempi kuin muilla vastaavaa rakennetta noudattavilla *nais*-alkuisilla yhdyssanoilla (vrt. esim. *naispappi*). Tämän tutkimuksen aineiston perusteella näyttää kuitenkin siltä, että vakiintuneiden yhdyssanojen kanssa homonyymisten uusien yhdyssanojen käyttö on melko yleistä. Homonyymit tarjoavatkin oivan keinon kielellä leikittelyyn. Homonyymien välttäminen vaikuttaa lähinnä yksittäisten yhdyssanojen leksikaalistumismahdollisuuksiin, ei niinkään kertamuodosteiden käyttöön tai rakenteiden produktiivisuuteen.

Toiseksi yhdyssanojen(kin) muodostusta rajoittavaksi leksikaaliseksi seikaksi voidaan laskea rakenteeseen sopivien lekseemien määrä kielessä. Johdosten kohdalla tämä tarkoittaa johdostyyppiin sopivien kantojen määrää kielessä (kannan merkitys ja morfologinen rakenne, ks. esim. Rintala 1978, Bauer 2001: 5–7). Yhdyssanojen kohdalla sopivien kantojen määrä on ensisijaisesti leksikaalissemanttinen rajoitus (esim. engl. *stepfather*, *-mother*, *?-grandparents*, *-aunt*, *-uncle*, *?-cousin*; Bauer 2001: 48). Jotkin yhdyssanarakenteet voivat kuitenkin vaatia myös osittain tiettyjen rakenteellisten ehtojen täyttymistä. Esimerkiksi suomen yhdyssanoissa yleinen *-s*-loppuinen yhdyssosamuoto voidaan muodostaa vain *-nen*-loppuisista sanoista, joten rakenteen yleisyys riippuu ainakin osittain *-nen*-loppuisten sanojen määrästä kielessä.

3.2.2 Prosessointiin liittyvät seikat

Kieliopillisten, semanttisten ja leksikaalisten rajoitusten lisäksi yhdyssanarakenteiden suosioon vaikuttavat niiden prosessoitiin liittyvät seikat. Kielellisen rakenteen prosessointiin kuuluu vuorovaikutuksessa kaksi erilaista osa-alueita: tuottaminen ja tulkitseminen. Helposti tuotettavat rakenteet eivät aina ole helppoja tulkita ja päinvastoin. Lisäksi esimerkiksi Dąbrowska (2008: 933) huomauttaa, että helpoimmin prosessoitavat rakenteet eivät aina ole esimerkiksi fyysisen tuottamisen kannalta taloudellisimpia. Kielellisten rakenteiden suosioon vaikuttaakin kaksi usein päinvastaista voimaa – pyrkimys helppouteen ja pyrkimys ymmärrettävyyteen (vrt. esim. ääntenmuutokset). Käytettävyydeltään optimaalinen on sellainen rakenne, jossa tuottamisen ja tulkitsemisen helppous ovat mahdollisimman hyvin tasapainossa.

Esimerkiksi Räisäsen (1986: 158) mukaan olennaista yhdyssanojen produktiivisuudelle on erityisesti nominatiivialkuisten yhdyssanan osien keskinäisen suhteen

semanttinen väljyys, jonka ansiosta yhdyssanoilla on lähes rajattomasti käyttömahdollisuuksia. Yhdyssanojen avulla voidaan helposti tiivistää laajoja asiakokonaisuuksia yhteen sanaan. Ne siis ovat sanojen muodostajan kannalta vaivattomia. Semanttinen väljyys saattaa kuitenkin tehdä uusista yhdyssanoista myös vaikeita ymmärtää (esim. otsikkokielen yhdyssanoista ks. Leskinen 1983).

Aina semanttinen väljyys ei edistä rakenteen produktiivisuutta (esim. Ryder 1994: 6; M. Jantunen 1997: 434). Yhdyssanojen tapauksessa semanttisesti väljän nominatiivialkuisen³⁵ rakenteen suurta suosiota (monissa kielissä tyyppi on yksinomainen, ja sisäisen taivutuksen mahdollisuutta pidetäänkin monissa kielissä yhdyssanuuden kriteerinä, ks. luku 2.1.3) voidaankin pitää osoituksena siitä, että väljyydestään huolimatta rakenne pystytään yleensä tulkitsemaan oikein. Viime kädessä oikean tulkinnan varmistaa konteksti. Toisin kuin monissa muissa kielissä, suomen kielessä viestin välittyminen voidaan yhdyssanoissa varmistaa (syntaktisten rakenteiden tapaan) myös alkuosan taivuttamisella (vrt. esim. sanoja *maahanmuuttaja* ja *maastamuuttaja* sanaan *?maamuuttaja*).

Kytömäen (1990: 71; ks. myös van Marle 1986: 608; toisaalta Plag 2003: 67–68, 2006: 552–553) mukaan sanaluokkaa muuttavilla johtimilla operointiin kuuluu kaksivaiheinen prosessi, jossa yritetään ensin löytää määrämerkityksinen johdin ja vasta sen puuttuessa valitaan yleismerkityksinen johdin. Yhdyssanoissa sekä rakenteiden produktiivisuus (ks. tarkemmin 8.1) että eräät teoriat sanatyypin synnystä (esim. Jackendoff 2009: 111–114) todistavat toiseen suuntaan toimivasta prosessista, jossa arvioidaan ensin, riittääkö nominatiivialkuinen rakenne varmistamaan viestin välittymisen, ja vasta sen ollessa epävarmaa valitaan toisenlainen, morfologisesti raskaampi rakenne.

Booijin (2007: 75) mukaan yhdyssanojen produktiivisuus on todennäköisesti pitkälti seurausta niiden semanttisesta läpinäkyvyydestä, joka helpottaa niiden ymmärtämistä. Vaikka yhdyssanarakenteessa osien keskinäinen semanttinen suhde on väljä, yhdyssanan osien merkitys on kielenkäyttäjille yleensä valmiiksi selvä. Näin pääteltäväksi jää ainoastaan osien välinen merkityssuhde. Johdosten yhteydessä myös morfologisen läpinäkyvyyden eli osien erottuvuuden on todettu edistävän rakenteen produktiivisuutta (esim. Hay–Baayen 2005: 345; M. Jantunen 1997: 434–435). Morfologinen läpinäkyvyys todennäköisesti edistää myös yhdyssanarakenteiden produktiivisuutta, ovathan yhdyssanan morfologiset osat yleensä hyvin selvästi erottuvia ja rakenteet helposti segmentoitavia.

Morfologisen ja semanttisen läpinäkyvyyden lisäksi rakenteen prosessointiin vaikuttaa myös rakenteen tuttuus kielenkäyttäjille. Mitä enemmän rakennetta käytetään, sitä tutumpi se on ja sitä helpommin prosessointi yleensä käy. Tutusta ja yleisestä rakenteesta on myös helppo ottaa mallia uusille muodosteille. Kielissä on havaittu diakroninen tendenssi, jossa ennestään suositut rakenteet kasvattavat suosiotaan ja harvinaiset rakenteet harvinaistuvat entisestään, kunnes mahdollisesti

35 Viitataan tässä nominatiivialkuisella myös muihin sitä semanttisesti vastaaviin rakenteisiin l. tietynlaisista sanoista mekaanisesti muodostettaviin alkuosan yhdyssanamuotoihin kuten *-nen*-loppuisen sanojen *-s*-loppuisiin yhdysosamuotoihin (esim. *hevostmies*).

menettävät produktiivisuutensa kokonaan. Yleensä kehitys ei kuitenkaan ole näin suoraviivaista, ja joskus jopa produktiivisuutensa lähes kokonaan menettänyt rakenne voi syystä tai toisesta nousta uuteen kukoistukseen. (Bauer 2001: 5–10; esim. sosiaalisista ja tyyllillisistä syistä ks. seuraava luku 3.2.3).

Rakenteen pääseminen sellaiseen suosioon, jossa se voi frekvenssinsä ansiosta kasvattaa suosiotaan entisestään, vaatii siltä alkujaankin tietynlaisia ominaisuuksia. Kielissä yleinen tendenssi on esimerkiksi lyhyiden morfologisten rakenteiden suosiminen pitkien sijaan (esim. M. Jantunen 1997: 435; syntaktisissa rakenteissa näin ei välttämättä ole, ks. esim. Helasvuo 2014: 31–34, 39–41). Myös morfologisesti yksinkertaiset rakenteet ovat yleisesti ottaen suositumpia kuin morfologisesti kompleksiset. Lyhyys ja morfologinen yksinkertaisuus liittyvät myös yhteen, sillä yleensä morfeemien lisääminen tekee sanoista myös pidempiä. Lyhyet ja yksinkertaiset rakenteet ovat helpompia prosessoida ja siten käytettävyydeltään parempia. Pitkiä ja kompleksisia sanoja on toisinaan moitittu myös suorastaan rumiksi (Kanervamäki 1996: tiivistelmä ja 18–20.) Yhdyssanoissa myös liiallinen tiivistäminen voi tosin aiheuttaa vaikeuksia tulkinnessa (Leskinen 1983).

Rakenteen yleisyyden lisäksi prosessointiin ja sitä kautta produktiivisuuteen vaikuttaa myös rakenteessa käytettyjen osien frekvenssi sekä sanaperheen koko (esim. Schreuder–Baayen 1997; Moscoso del Prado Martín ym. 2004). Sanaperheen koolla (*family size*) tarkoitetaan useimmiten tietystä sanasta muodostettujen johdosten määrää. Perhetaajuus (*family frequency*) puolestaan on kaikkien tietystä sanasta muodostettujen johdosten esiintymien summa esimerkiksi korpuksessa. Molempia käsitteitä käytetään toisinaan myös laajemmassa merkityksessä, jolloin sanaperheeseen lasketaan myös yhdyssanat, joiden osana sana esiintyy. Suomen kaltaisissa kielissä, joissa yhdyssanat ovat erittäin produktiivinen luokka, tämän näkökulman ongelma on kuitenkin leksikkoon laskettavien yhdyssanojen erottaminen niin sanotuista potentiaalisista sanoista (Nikolaev – Jussi Niemi 2005: 487; Nikolaev 2011: 42–45).

Sanaperheen suuren koon on todettu nopeuttavan sanojen prosessointia (esim. Schreuder–Baayen 1997). Esimerkiksi suomen kielessä monien sanojen perhekoko on niin suuri, että mahdollisesti ainoastaan semanttisesti relevantit jäsenet vaikuttavat sanojen prosessointiin. Lisäksi perheeseen kuuluvilla yhdyssanoilla ja johdoksilla voi olla erilainen vaikutus prosessointiin. Asiaa ei kuitenkaan ole toistaiseksi tutkittu. (Nikolaev 2011: 44–45.) Perhetaajuudella puolestaan on tutkimuksissa havaittu olevan päinvastaisia vaikutuksia prosessointiin kuin perhekoolla: korkea perhetaajuus hidastaa sanojen prosessointia (esim. Baayen–Tweedie–Schreuder 2002; Pykkänen–Feintuch–Hopkins–Marantz 2004). Tosin joissakin tutkimuksissa perhetaajuudella ei havaittu lainkaan vaikutusta prosessointiin (esim. Schreuder–Baayen 1997; de Jong–Schreuder–Baayen 2000). Tässä tutkimuksessa en tarkastele perhekoon ja perhetaajuuden kaltaisia yksittäisiin leksemeihin sidoksissa olevia seikkoja. Luvussa 6.2 käsittelem kuitenkin eri yhdyssanarakenteiden leksikaalista monipuolisuutta tarkoituksenani osoittaa, että joissakin rakenteissa leksikaalinen variaatio on laajempaa kuin toisissa.

Sanojen prosessointiin vaikuttaa myös niiden leksikaalistuneisuuden aste (leksikaalistumisesta ks. s. 31–32 luvussa 2.1.3). Leksikaalistuneet, vakiintuneet ja tilapäiset yhdyssanat prosessoidaan eri tavalla: leksikaalistuneet kokonaisuuden kautta (tosin näyttöä jopa opaakkien yhdyssanojen jakamisesta osiin on saatu, ks. Jarema ym. 1999; Ulfsbjorninn 2009; luku 2.3.3), tilapäiset yhdyssanat osiansa kautta, vakiintuneet joko osiansa tai kokonaisuuden kautta sen mukaan kumpi on nopeampaa (ks. myös luku 2.3.3). Johdostyyppien tutkimuksessa on havaittu, että mitä enemmän tyyppiin kuuluu leksikaalistuneita johdoksia, sitä vähemmän tyyppi tuottaa uusia sanoja (Bybee 1985: 18; Baayen–Lieber 1991: 804–805). Fraseologian ja syntaktisten rakenteiden tutkimuksessa puolestaan on huomattu, että leksikaalistumat päinvastoin edistävät tyyppin produktiivisuutta, sillä ne toimivat analogisina malleina uusille muodosteille (esim. Heinonen 2010: 364). Yhdyssanat ovat leksikaalistumisen ja produktiivisuuden suhteen sikäli syntaktisten rakenteiden kaltaisia, että produktiivisiinkin tyyppeihin kuuluu paljon leksikaalistuneita muodosteita. Tämä voi johtua yhdyssanojen ja syntaktisten rakenteiden johdoksia laajemmasta produktiivisuudesta. Toisaalta johdosten kanssa yhdyssanoilla on yhteistä se, että mitä epäproduktiivisempi rakenne on, sitä suurempi suhteellinen osuus sitä noudattavista yhdyssanoista todennäköisesti on leksikaalistuneita (ks. esim. yhdyssanan osien kongruenssia käsittelevä luku 5.3.3).

3.2.3 Sosiaaliset ja tyylilliset seikat

Rakenteen omat ominaisuudet ja niiden prosessointiin liittyvät seikat eivät yleensä yksin riitä selittämään rakenteen suosiota. Vaikutusta on myös erilaisilla sosiaalisilla ja tyylillisillä seikoilla. Niitä on kuitenkin tutkittu produktiivisuuden yhteydessä toistaiseksi vasta vähän. Plagin (2006: 550) mukaan yksi ilmeisimmistä rakenteen käytännön produktiivisuuteen vaikuttavista kielenulkoisista seikoista on muoti. Kielessä voivat olla muotia sekä yksittäiset muotisanat että erilaiset rakenteet ja sananmuodostustavat. Ainakin yksittäiset ilmaukset voivat tulla muotiin yhden tarpeeksi vaikutusvaltaisen puhujan tai vaikkapa suosituksen tv-sarjan vaikutuksesta. Muoti-ilmaukset voivat kulkeutua myös kielirajojen yli, ja osa kielen muoti-ilmiöistä onkin kansainvälisiä. (Iisa–Oittinen–Piehl 2002: 215–216.) Muiden muoti-ilmiöiden tapaan myös kielelliset ilmiöt voivat olla muodikkaita vain jonkin ihmisryhmän keskuudessa (esim. ikäryhmät). Kielellisen muoti-ilmiöt voivat rajoittua myös tiettyihin kielimuotoihin.

Kielellisten muotien synnyn ja keston ennustaminen on vaikeaa. Ei ole myöskään täysin selvää, milloin ilmaisu tai rakenne on katsottava muodiksi ja milloin vain muuten kielessä yleiseksi. Iisan, Oittisen ja Piehlin (2002: 216) mukaan erilaisille kielellisille muodeille on yhteistä se, että ne tekevät ilmaisusta epätasällistä. Vaikka ilmauksessa sinänsä ei olisi mitään vikaa, se muuttuu epätasälliseksi levitessään sellaisiin käyttöympäristöihin, joihin se ei oikeastaan sovi. Muoti-ilmaus voi levitessään myös syrjäyttää aiemmin käytössä olleita tasällisiä ja havainnollisia ilmaisutapoja. Myös Romalo (1992) mainitsee kielellisen muodin tunnusmerkiksi ilmauksen käytön laajan leviämisen, frekvenssin nopean kasvun ja siitä johtuvan

muiden vastaavassa käytössä aiemmin olleiden ilmausten käytön vähenemisen sekä siitä seuraavan kommunikaation köyhtymisen. Lisäksi Romalo mainitsee kielellisen muoti-ilmiön tunnusmerkiksi kritiikittömän ja matkivan käytön.

Kielen muoteja on tutkimuksessa ja kielenhuollossa käsitelty ensisijaisesti sanojen ja rakenteiden valinnan (esim. Romalo 1992) sekä nimityksen näkökulmasta (esim. Arjava 2005: esim. 264–269; Kiviniemi 2006: 168–226; Kryukova 2008). Fennistiikassa nimenomaan yhdyssanojen kohdalla on kiinnitetty huomiota myös tiettyjä leksikaalisia osia hyödyntävien sanojen muodostamiseen (esim. *-vapaa-* ja *-ystävällinen-*loppuiset yhdyssadjektiivit, mm. Vesikansa 1989: 241–250; Palander 2002). Myös turhan pitkiä yhdyssanoja on pidetty muoti-ilmiönä (esim. Iisa–Oittinen–Piehl 2002: 226–227).

Tämän tutkimuksen kohteena olevassa toimitetussa lehtikielessä muoti-ilmausten liikaan käyttöön puututaan yleensä melko nopeasti. Ammattitaitoinen toimittaja välttää epätasällisiä ja kenties lukijakuntaa ärsyttäviä muoti-ilmauksia. Vihjeitä vältettävistä ilmauksista toimittajat saavat esimerkiksi kielenhuollolta. Kielenhuollon suosituksilla todennäköisesti onkin (tai ainakin toivotaan olevan) useimmiten vaikutusta rakenteiden ja ilmaisutapojen suosioon. Yhdyssanojen kohdalla esimerkiksi E. A. Saarimaan antamat ohjeet yhdyssanan alkuosan muodosta ovat todennäköisesti lisänneet genetiivialkusten yhdyssanojen suosiota (Kettunen 1949: 103–113; Vesikansa 1989: 228; T. Kolehmainen 2011, 2014: 286–299).

Kielenhuolto yritti pitkään vastustaa varsinkin niin sanottuja takaperoisesti johdettuja³⁶ yhdyssverbejä (esim. Saarimaa 1971 [1947]: 88–89), mutta myös niiden tarpeellisuudelle löytyi puolustajia jo 1980-luvulla (esim. Rahtu 1984; Vesikansa 1989: 257–258). Yhdyssverbien tarpeellisuudesta kielii myös se, että niiden määrän on arveltu olevan kasvussa (esim. ISK:n [§ 430–431] mukaan yhdyssverbit ovat karttuva ryhmä). Lopullisen päätöksen rakenteen tarpeellisuudesta ja käytöstä tekevätkin tavalliset kielenkäyttäjät, ei kielenhuolto tai muodinluojat.

Kielenkäyttäjät tekevät lopullisen³⁷ päätöksen myös sanojen ja rakenteiden sopivuudesta eri kielimuotoihin. Tämä on olennaista myös produktiivisuuden kannalta, sillä rakenteen sopiminen mahdollisimman moniin eri kielimuotoihin ja rekistereihin edistää sen produktiivisuutta (Baayen 1992: 109–110). Lisäksi produktiivisuutta edistää myös rakenteen sopiminen mahdollisimman monenlaisten kielenkäyttäjien suuhun. Esimerkiksi Tanja Säily (2008, 2014) on havainnut myös sosiologisten taustatekijöiden kuten iän, sukupuolen, koulutustason jne. voivan vaikuttaa rakenteen produktiivisuuteen. Yhdyssanojen kohdalla esimerkiksi viittauskohteen

36 Ns. takaperoisesti johdetuissa yhdyssverbeissä edusosa on johdettu nominista, joka puolestaan on johdettu verbistä (esim. *vakuuttaa* > *vakuutus* mutta *palovakuutus* > *palovakuuttaa*). Ns. normaalijohdoisissa yhdyssverbeissä puolestaan edusosan johtosuunta on sama kuin kantasanassakin (esim. *kuva* > *kuvata* ja *valokuva* > *valokuvata*). (Ks. esim. Rahtu 1984: 413–418; Vesikansa 1989: 256–258). Normaalijohdoisten ja takaperoisesti johdettujen yhdyssverbien erottamista toisistaan on pidetty vaikeana jopa kieliasiantuntijoille saati maallikoille (Rahtu 1984: 422–423).

37 Lopullisella tarkoitan tässä päätöstä siitä, missä kielimuodoissa sanoja ja rakenteita käytetään tai missä niiden käyttö katsotaan yleisesti sopivaksi. En siis tarkoita lopullisella tässä muuttumatonta; sanojen ja rakenteiden tyyliarvohan saattaa muuttua nopeastikin (esim. Grönros 1994).

prototyypillisyyttä korostavat toistoyhdyssanat (esim. *ruokaruoka*, *kotikoti*) kuuluvat ainakin aluksi ensisijaisesti nuorten kieleen. Hyvä on huomata myös, että kielentutkijan koulutuksen saaneiden intuitio rakenteen produktiivisuudesta ei aina välttämättä vastaa koko kieliyhteisön todellisuutta, sillä esimerkiksi tietoisessa sanastotyössä voidaan käyttää sellaisiakin rakenteita, joita tavalliset kielenkäyttäjät eivät käytä (esim. *-in*-tekimennimijohdin, Rintala 1978: 154). Sosiaalisten taustatekijöiden lisäksi vaikutusta on tietysti myös jokaisen kielenkäyttäjän idiolektillä ja esimerkiksi asenteella sananmuodostusta kohtaan: kielestä kiinnostunut leikkelijä muodostaa varmasti erilaista uutta sanastoa kuin ihminen, joka ei leikkittelystä tai kielestä juuri perusta.

3.3 Produktiivisuuden mittaaminen

Tässä luvussa esittelen neljä produktiivisuuden mittaamiseen soveltuvaa metodologiaa, joista viimeinen on tietääkseni ensimmäistä kertaa tällaisenaan käytössä tässä tutkimuksessa. Produktiivisuuden mittaamisessa käytettyjen metodien sopivuus ja luotettavuus vaihtelevat sen mukaan, miten produktiivisuus on tutkimuksessa määriteltä ja mitä ilmiön eri puolista tahdotaan tarkastella. Myös saatavissa oleva aineisto vaikuttaa olennaisesti metodin valintaan.

Produktiivisuuden tutkimuksessa on aiemmin käytetty pääasiassa kolmea erilaista metodologiaa. Kaikkia kolmea metodologiaa kohtaan on esitetty kritiikkiä, ja useat tutkijat ovatkin todenneet, että mahdollisimman luotettavien tulosten saamiseksi produktiivisuuden mittaamisessa olisi parasta käyttää rinnakkain useita metodeja (Bolzky 1999: 191–194; Plag 1999: 117; Bauer 2001: 158–161). Monimenelelmäisistä tutkimuksista onkin saatu lupaavia tuloksia (esim. Bolzky 1999; Plag 1999). Tämän tutkimuksen tarpeisiin aiemmat metodit eivät kuitenkaan sovi edes rinnakkain käytettynä, sillä niissä aineistona käytettävät sanakirjat ja suuret tekstikorpuksot ovat suomen kielen yhdyssanojen kannalta ongelmallisia: läpinäkyviä yhdyssanoja ei yleensä oteta sanakirjoihin ja laajojen tekstikorpuksien automaattinen prosessointi ei ainakaan toistaiseksi ole suomen kielen yhdyssanojen osalta riittävän luotettavaa. Myös muiden kielen rakenteiden kuin yhdyssanarakenteiden tutkimuksessa on tilausta uudentilaisille produktiivisuuden mittaamenetelmille (esim. Bolzky 1999: 191–194). Tulevaisuudessa tässä tutkimuksessa käytettyä metodologiaa voidaan toivottavasti käyttää myös muiden rakenteiden tutkimuksessa.

Alaluvuissa 3.3.1–3.3.3 esittelen ensin aiemmin produktiivisuuden tutkimuksessa käytetyt metodit eli 1) eri-ikäisten sanastojen vertailuun perustuvan metodin (3.3.1), 2) tutkittavaa rakennetta noudattavien korpuksissa esiintyvien kertamuodosteiden määrään perustuvan metodin (ns. Baayenin metodi, 3.3.2) sekä 3) kielenkäyttäjillä teetettäviä tuottamis- ja päätöksentekotehtäviä hyödyntävän metodin (3.3.3). Pohdin kunkin menetelmän hyviä ja huonoja puolia sekä sopivuutta erityisesti yhdyssanarakenteiden produktiivisuuden mittaamiseen.

Neljännessä alaluvussa (3.3.4) esittelen tässä tutkimuksessa käytetyn metodin (katson metodin esittelyn liittyvän olennaisesti tutkimuksen teoreettiseen taustaan; tutkimuksen käytännön menetelmät on esitelty luvussa 4). Metodi perustuu rakenteiden tyyppifrekvenssin tarkasteluun nimenomaan uudessa sanastossa. Analyysin toinen askel on varsinkin kvantitatiivisessa analyysissä harvinaisiksi tai muuten poikkeavaksi osoittautuneiden rakenteiden kvalitatiivinen tarkastelu. Yksi tutkimukseni tarkoituksista on testata uutta metodologiaa ja arvioida sen sopivuutta produktiivisuuden mittaamiseen. Metodien käytölle on edellytyksenä sopivan aineiston saatavuus. Suomessa metodin käytön mahdollistaa Kotimaisten kielten keskuksessa koottava Nykysuomen sanastotietokanta (NSST; ks. 4.1).

3.3.1 Eri-ikäisten sanastojen vertailu

Esimerkiksi Shmuel Bolozky (1999: 5, 194; ks. myös Plag 1999 96–105, 117–118; Haarala–Nissinen 1994: 23–24; Bauer 2001: 156–161; Grönros 2002: 8) ehdottaa morfologisen produktiivisuuden mittaamiseen – muiden menetelmien ohella – lähestymistapaa, jossa verrataan tiettyä rakennetta noudattavien vakiintuneiden ja uusien sanojen määrää kielessä. Käytännössä tämä tehdään poimimalla ensin kahdesta eri-ikäisestä sanastosta (yleensä sanakirjasta) tutkittavaa rakennetta noudattavat sanat ja vertaamalla sitten tietoja keskenään (käytännön toteutusta esittelee tarkemmin esim. Iikkanen 1993). Rakenteen produktiivisuuden asteena voidaan pitää vain uudessa aineistossa esiintyvien muodosteiden osuutta. Laurie Bauerin (2001: 156–158) mukaan metodin avulla voidaan tarkastella erityisesti produktiivisuuden muutoksia. Kritiikkiä metodologiaa kohtaan ovat esittäneet esimerkiksi Baayen ja Renouf (1996: 69).

Produktiivisuuden tutkimuksessa elää myytti, jonka mukaan aiemmissä tutkimuksissa käytettiin usein tutkimusmenetelmää, jossa produktiivisuutta mitattiin yksinkertaisesti laskemalla tutkittavaa rakennetta noudattavat sanat kielen sanastosta (Bauer 2001: 144). Tällaisen (yleiseksi luullun) tarkastelutavan ongelma on, että se ei ota huomioon rakenteen produktiivisuudessa ajan kuluessa mahdollisesti tapahtuneita muutoksia: se mikä oli produktiivista joskus aiemmin, ei välttämättä ole produktiivista enää nykyään (Rintala 1978: 155; Baayen 1992: 110–111; Bauer 2001: 6–9; ks. myös luku 3.1.2). Sen sijaan eri-ikäisten sanastojen vertailuun perustuvassa metodissa jo päättyneen ja yhä jatkuvan produktiivisuuden ero otetaan huomioon rajaamalla aineisto sanojen oletetun syntyhetken mukaan.

Eri-ikäisten aineistojen vertailuun perustuvaa metodologiaa hyödyntävien tutkimusten suurimmat ongelmat liittyvät muodosteiden iän tunnistamiseen sekä aineiston valintaan ja saatavuuteen. Menetelmässä yleisimmin käytetty aineisto on eri-ikäiset sanakirjat sekä uudissanakirjat ja sanakirjojen täydennysosat. Periaatteessa aineistona voidaan käyttää myös esimerkiksi tekstikorpuksia. Käytännössä sanakirjat ovat kuitenkin kattavampia aineistoja, sillä suurissakin tekstikorpuksissa on yleensä vähemmän eri lekseemejä kuin sanakirjoissa, ellei kyseessä ole valtava korpus tai hyvin pieni sanakirja. (Bauer 2001: 158–161.) Mikäli aineistoksi valitaan laajat tekstikorpukset, ratkaistavaksi tulee myös aineiston automaattiseen (esi)käsitteilyyn

liittyviä ongelmia (yhdyssanojen automaattisen analysoinnin ongelmista ks. tarkemmin seuraava alaluku 3.3.2).

Uusien sanojen tunnistaminen vain niiden sanakirjaesiintymien perusteella on epävarmaa, sillä sanakirjojen ei ole yleensä tarkoitukseen kuvata kielen kaikkea sanastoa. Varsinkin monet hyvin produktiivisina ja läpinäkyvinä pidettyjä rakenteita noudattavat sanat jätetään tietoisesti sanakirjojen ulkopuolelle, sillä niiden ajatellaan olevan ymmärrettäviä pelkän kielitiedon perusteella. Plagin (1999: 97–100, 243–244; ks. myös Berg 1991: 4) mukaan Oxford English Dictionary (OED 1884–) sopii monista muista sanakirjoista poiketen produktiivisuuden muutoksia käsittelevän tutkimuksen aineistoksi, sillä sen tavoitteena on kielen mahdollisimman täydellinen kuvaaminen ja siihen otetaan siksi myös läpinäkyvät sanat. OED:ssä ilmoitetaan lisäksi tiedot sanan ensimmäisestä atestaatiosta, joten myös sanojen iän arviointi on sen avulla muita sanakirjoja luotettavampaa. Kuitenkaan harvoista kielistä on saatavilla sopivia aineistoja (Bauer 2001: 158). Laajimmassa suomen yleiskieltä kuvaavassa sanakirjassa eli Nykysuomen sanakirjassa (NS 1951–1961) hakusanoja on noin 201 000, kun OED:ssä niitä on noin 600 000³⁸. Sekä Suomen kielen perussanakirjassa (PS 1990–1994) että Kielitoimiston sanakirjassa (KS 2004–) hakusanoja on noin 100 000³⁹. Vertailtavaksi olisi parasta valita keskenään samantyyppiset ja samankokoiset aineistot, joten suomen kannalta ongelmallista on myös se, että yleiskielen sanakirjamme ovat keskenään erikokoisia ja osin eri kriteerein koostettuja (ks. esim. Grönros 2007: 15). Niissä ei myöskään ole tietoja esimerkiksi sanojen ensiesiintymistä.

Puutteista huolimatta monet tutkijat pitävät OED:n ohella myös muita sanakirjoja hyvänä lähteenä morfologisen produktiivisuuden tutkimuksessa, kunhan niiden vääristymät otetaan huomioon (Bolzky 1999: 194; Plag 1999: 243–244; Bauer 2001: 161). Myös Suomessa Risto Haarala ja Irma Nissinen (1994: 23) sekä Eija-Riitta Grönros (2002: 8) ovat ehdottaneet yleiskielen sanakirjojemme käyttöä produktiivisuuden muutosten ja suomen uudissanakerrostumien tarkastelussa. Tosin myös Grönros (mp.) huomauttaa, että uudissanonjen rajaaminen sanakirjojen perusteella ei ole ongelmatonta.

Yksi sanakirja-aineiston kenties suurimmista vääristymistä koskee yhdyssanoja, sillä usein nimenomaan läpinäkyvinä pidetyt yhdyssanat ovat niitä, jotka jätetään aineistoa karsittaessa sanakirjan ulkopuolelle (ks. esim. NS: Alkusanat: V; Grönros 1991: 33; Haarala–Nissinen 1994: 24–25). Useimpiin sanakirjoihin ei myöskään

38 OED online -sivuston mukaan sanakirja kokoa yhteensä tiedot 600 000 sanan merkityksestä, historiasta ja ääntämyksestä (OED: About). Toisaalla sivustolla todetaan kuitenkin, että tällä hetkellä OED sisältää noin 270 000 päähakusanaa ja 220 000 alihakusanaa (OED: Glossary).

39 NS:n hakusanoista noin 65 % on eri-ikäisiä yhdyssanoja (Saukkonen 1973: 338). PS:n noin 20 000:sta NS:aan verrattuna uudesta sanasta arviolta viidesosa on eli noin 4 000 kpl yhdyssanoja (Haarala–Nissinen 1994: 24). KS:ssa vuosien 2004 ja 2006 painoksissa puolestaan on noin 6 000 PS:aan verrattuna uutta hakusanaa (Grönros 2004: 4, 2006: 31; KKK: Sanakirjoja). Tietoa yhdyssanojen määrästä tässä joukossa ei ole. Tämän tutkimuksen tutkimusaineistossa 28 091 uutta yhdyssanaa (ks. aineiston tarkempi esittely luvusta 4). Sanakirjoja vertailemalla saatavissa olevaan aineistoon verrattuna tämän tutkimuksen aineisto on siis varsin kattava.

oteta kertamuodosteita tai muita harvinaisuuksia. (Baayen–Renouf 1996; Bauer 2001: 158; Grönros 2002: 8–10.) Suomen yleiskielessä käytettävistä yhdyssanoista kuitenkin todennäköisesti jopa yli puolet on kertamuodosteita (Mäkisalo 2000a: 14). Sanakirja-aineiston käyttö nimenomaan yhdyssanarakenteita koskevassa tutkimuksessa olisikin vähintään erikoinen ratkaisu.

3.3.2 Kertamuodosteiden määrä korpuksessa (ns. Baayenin metodi)

Viime vuosikymmeninä yksi suosituimmista tavoista morfologisen produktiivisuuden mittaamiseen on ollut alun perin Harald Baayenin (1989, 1992, 1993; Baayen–Lieber 1992) esittelemä laskentamalli, joka perustuu laajassa tekstikorpuksessa esiintyvien tiettyä rakennetta noudattavien kertamuodosteiden määrään. Lyhyiden vuoksi kutsun menetelmää tässä luvussa Baayenin metodiksi. Matti Räsänen (2010: 205) pukee Baayenin metodin perusperiaatteen sanoiksi yksinkertaisesti: ”Produktiivisuuden arvioinnin yleisperiaate on aineistosta riippumatta selvä: mitä enemmän konstruktio sisältää kerrallisia ilmauksia, sitä produktiivisempi se on.”

Baayenin metodi perustuu havaintoon siitä, että epäproduktiivisen johtimen sisältävät sanat esiintyvät kielenkäytössä yleensä frekventimmin kuin produktiivisen johtimen sisältävät sanat (produktiivisuuden, rakennetta noudattavien sanojen frekvenssin ja rakenteen iän suhteeseen on kiinnittänyt huomiota jo esim. Laalo 1982). Ilmiö selittyy sekä frekvenssiin että produktiivisuuteen yhteydessä olevan semanttisen kompleksisuuden kautta: Epäproduktiivisen johtimen sisältävät sanat ovat leksikaalistuneita eli tiettyyn (erikoistuneeseen) merkitykseen tai useisiin merkityksiin vakiintuneita. Ne ovat siis kiteytyneitä ja todennäköisesti sellaisenaan muistiin painettuja. Sen sijaan produktiivisten prosessien kautta syntyneiden tilapäismuodosteilla on vain yksi (vakiintumaton) merkitys, joka määräytyy ainoastaan niiden rakenteen perusteella. Epäproduktiivisen johtimen sisältävät sanat ovat siis semanttisesti kompleksisempia. Semanttisesti yksinkertaisempien sanojen on havaittu esiintyvän kielessä frekventimmin kuin semanttisesti kompleksisten. Leksisiin frekvenssiä voidaan siis käyttää sen semanttisen kompleksisuuden epäsuorana mittarina. Koska semanttinen kompleksisuus puolestaan heijastaa produktiivisuutta, leksisiin frekvenssiä voidaan käyttää myös produktiivisuuden epäsuorana mittarina. Näin ollen voidaan olettaa, että mitä enemmän (riittävän suuressa) korpuksessa on tiettyä rakennetta noudattavia kertamuodosteita, sitä produktiivisempi rakenne on⁴⁰. (Aronoff 1983; Baayen–Lieber 1991: 805, 813; Bauer 2001: 147–155.) Tämä ei tarkoita, että korpuksessa esiintyvät kertamuodosteet välttämättä olisivat uudismuodosteita. Kertamuodosteiden määrä kuitenkin heijastaa hyvin rakenteelta odotettavissa olevien uusien muodosteiden määrää (Baayen–Lieber 1991: 813).

Käytännössä rakenteen produktiivisuusasteen mittaaminen tehdään Baayenin metodissa tilastollisten tunnuslukujen avulla. Eri tunnuslukuja on kolme. Niiden

40 Tiettyä rakennetta noudattavien leksiemien esiintymistiheyttä ei kuitenkaan voida pitää produktiivisuuden suorana mittana, sillä semanttinen kompleksisuus koskee myös perussanoja (Bauer 2001: 147).

kaikkien taustalla on sama edellä esitelty peruseriaate, mutta ne ottavat eri tavoin huomioon esimerkiksi käytetyn korpuksen koon (tunnuslukujen P , P^* ja P^* las-kentakaavoista ja merkityksistä ks. tarkemmin Baayen 1989, 1992, 1993; Baayen–Lieber 1991). Tunnusluvut voidaan nähdä metodin eri versioina ja jatkokehittelynä (Bauer 2001: 147–156), mutta Baayenin (1993: 205) itsensä mukaan tunnuslukuja voidaan käyttää myös rinnakkain.

Baayenin metodologiaa kohtaan on esitetty sekä sen teoreettiseen taustaan että käytännön toteutukseen ja käytettyihin tunnuslukuihin kohdistuvaa kritiikkiä (esim. van Marle 1992; Plag 1999: 112–113; Bauer 2001: 147–160). Yksi menetelmän kärkkäimmistä kritisoijista on ollut Jaap van Marle (1992: 156), joka on kyseenalaistanut koko metodin teoreettisen taustan. Hänen mukaansa uudismuodosteiden rakenteen ja aiemmin syntyneiden samaa rakennetta noudattavien sanojen esiintymisfrekvenssillä ei ole minkäänlaista suoraa yhteyttä toisiinsa. Metodin ongelma onkin se, että suoraa tai edes epäsuoraa yhteyttä kertamuodosteiden ja uudismuodosteiden välillä ei voida todistaa. Lisäksi vakiintuneet mutta harvinaiset sanat voivat esiintyä laajassakin korpuksessa vain kerran, tai sanan kertaesiintyminen voi johtua silkasta sattumasta (Bauer 2001: 147; Pustyl'nikov–Schneider–Wiejowski 2009: 15). Toisaalta uusikin muoti-ilmaus tai nopeasti kieleen vakiintunut muodoste voi esiintyä korpuksessa useita kertoja.

Baayen ei esitä selvää vastausta van Marlen (1992: 156) kritiikkiin (ks. kuitenkin Baayen 1993: 188). Tästä ja muusta arvostelusta huolimatta Baayenin metodin eri kehitysvaiheet ovat olleet tärkeitä edistysaskeleita todellisen kielenkäytön sananmuodostusprosessien ja niiden produktiivisuuden tutkimuksessa. Esimerkiksi Plag (1999: 111–115, 244) toteaa eri produktiivisuudenmittaustapoja vertailevassa ja yhdistelevässä tutkimuksessaan, että metodi on käyttökelpoinen, kunhan sen rajoitukset otetaan huomioon ja täydennetään analyysia myös kvalitatiivisilla huomioidilla (ks. myös Bauer 2001: 150–151). Samoilla linjoilla on myös Bolozky (1999: 191–192), joka korostaa toisiaan täydentävien metodien käyttöä produktiivisuuden tutkimuksessa ylipäänsä.

Suomen kielen tutkimuksessa on käytetty Baayenin metodologiaa vähän (ks. kuitenkin Nikolaev – Jussi Niemi 2008; kertamuodosteiden määrään kiinnittää huomiota myös Räsänen 2010). Mäkisalo (2000a: 14) on testannut metodologiaa suomen kielen yhdyssanoihin ylipäänsä ja todennut, että yhdistäminen todella on suomessa hyvin produktiivinen sananmuodostuskeino: Turun Sanomista kootun suuren sanomalehtikorpuksen (M. Laine – Virtanen 1996) aineistossa 58 % yhdyssanoista on kertamuodosteita. Eri yhdyssanarakenteiden produktiivisuuden mittaamiseen Baayenin metodologiaa ei ole sovellettu. Mäkisaloon (mp.) mukaan metodi ei välttämättä edes sovellu yhdyssanarakenteiden tutkimukseen yhtä hyvin kuin johdostyyppien tutkimukseen.

Baayenin metodin käyttö suomen kielen yhdyssanarakenteiden tutkimuksessa on vaikeaa myös käytännön syistä. Methodiin kuuluvat olennaisena osana laajat tekstikorpuksot, jotka on käytännössä välttämätöntä esianalysoida koneellisesti. Varsinkin alhaisen frekvenssin sanojen oikea analysointi on koneellisin menetelmin usein

epävarmaa, mutta kertamuodosteiden määrään perustuvassa mallissa juuri alhaisen frekvenssin sanojen oikea analysointi on keskeistä oikean tuloksen saamiseksi (Evert–Lüdeling 2001). Automaattisen prosessoinnin ongelmat korostuvat suomen kaltaisissa morfologisesti rikkaissa kielissä, joissa on paljon sananmuotohomonymiaa, joka vaikeuttaa toimivien työkalujen kehittämistä. Produktiivisuutta koskevassa tutkimuksessa on hyvä muistaa myös se, että nimenomaan uudissanojen analysointi koneellisin menetelmin on epävarmaa.

Ainakaan toistaiseksi varsinkaan suomen kielen yhdyssanojen sisäisen rakenteen automaattinen prosessointi ei ole vielä riittävän luotettavaa Baayenin metodin soveltamiseen yhdyssanarakenteiden produktiivisuuden tutkimuksessa. Esimerkiksi Krister Lindénin ja Tommi Pirisen (2009) kehittämässä suomen kielen yhdyssanojen prosessointiin (erityisesti segmentointiin) sopivassa HFST-LexC-työkalussa useat harvinaiset yhdyssanarakenteet on jätetty tarkoituksella huomiotta. Myös Omorfijäsentimen tätä tutkimusta tehdessä käytettävissä olleissa versioissa on puutteita nimenomaan uusien ja vakiintumattomien yhdyssanojen sisäisen rakenteen analysoinnissa⁴¹ (Omorfista ks. tarkemmin esim. Pirinen 2011; Omorfi 2014).⁴²

3.3.3 Kielenkäyttäjillä teetettävät tuottamis- ja päätöksentekotestit

Kolmas produktiivisuuden mittaamisessa käytetty metodi on erilaiset kielenkäyttäjillä teetettävät tuottamis- ja päätöksentekotestit, joissa kielenkäyttäjää pyydetään joko itse tuottamaan uusia sanoja (tai sanamuotoja) tai esittämään arvioita tutkijan tutkimusta varten tuottamista uusista sanoista (tai sanamuodoista). Usein testeissä mitataan myös koehenkilön reaktioaika. Esimerkiksi EEG-mittauksilla ja silmäliikekameroilla voidaan mitata myös koehenkilön muita reaktioita. Tuottamis- ja päätöksentekotestejä käytetään produktiivisuuden tutkimuksen lisäksi myös kielen prosessoinnin ja oppimisen tutkimuksissa. (Ks. esim. Bertram–Hyönä 2003, 2013; Boudelaa–Marslen-Wilson 2011; Pollatsek–Bertram–Hyönä 2011; Havas–Rodríguez–Fornells–Clahsen 2012).

Kielenkäyttäjille tehtävät testit keskittyvät nimenomaan sanojen tuottamisprosessiin ja kieleen tarjolla olevien sanojen sopivuuden arviointiin (vrt. muut tässä luvussa esitellyt metodit, ks. luvut 3.3.1, 3.3.2 ja 3.3.4). Niiden avulla ei siis yleensä tutkita kielessä jo käytössä olevia sanoja. Menetelmässä keskeinen käsite on potentiaalinen sana (käsitteestä tarkemmin Coseriu 1971, 1975; Aronoff 1976; Aronoff–Schvaneveldt 1978: 108–109; Plag 1999: 6–11). Potentiaalinen sana on mahdollinen mutta ei vielä käytössä esiintynyt sana, jolle ei ole vakiintunut muuta merkitystä kuin sen osien perusteella pääteltävissä oleva merkitys (luku 2.3.1). Kielenkäyttäjien reaktioita testaamalla potentiaaliset sanat voidaan asettaa järjestykseen sen

41 Tieto perustuu syksyllä 2012 Omorfin kehittäjän Tommi Pirisen kanssa käytyyn keskusteluun sekä omiin kokeiluihini Omorfin eri versioilla.

42 Vuoden 2014 lopulla tämän tutkimuksen analyysien ollessa jo valmiita ensisijaisesti oikolukuun kehitettyjen Voikko-työkalujen (Voikko) yhdyssanojen prosessointi on jo huomattavan edistynyt. Voikon verkkosivujen (Voikko: UKK) mukaan Voikossa on haluttu kiinnittää erityistä huomiota yhdyssanojen prosessointiin.

mukaan, kuinka todennäköisesti niitä tullaan kielessä käyttämään (Aronoff–Schvaneveldt 1978: 108–109). Koska yksittäisen sanan ilmaantuminen kieleen riippuu ensisijassa kielenulkoisen maailman tapahtumien synnyttämästä tarpeesta sanan käytölle, lekseemitasoa järkevempi tarkastelutaso on rakenteiden aktuaalistumistodennäköisyys uudessa sanastossa.

Yksittäisistä produktiivisuuden mittaamisessa käytetyistä testeistä tunnetuin lienee Jean Berkon (1958) kehittämä *wug*-testi, jonka avulla on alkujaan tutkittu lasten kielenoppimista. Testiä on myöhemmin käytetty myös taivutusmorfologian produktiivisuuden tutkimuksessa (suomen osalta esim. Nikolaev – Jussi Niemi 2006). *Wug*-testissä koehenkilöitä pyydetään taivuttamaan epäsanuja. Näin saadaan tietoa esimerkiksi siitä, mitä taivutustyyppisiä kielenkäyttäjät suosivat ja millä rakenteellisilla ehdoilla kielenkäyttäjät sitovat uusia sanoja taivutusluokkiin.

Johtomorfologian produktiivisuutta koskevissa tutkimuksissa yleisemmin käytettyjä ovat aukkotehtävät, joissa koehenkilön tulee täydentää lause joko oman mielensä mukaan tai johtamalla aukkoon sopiva sana hänelle valmiiksi annetuista kannasta (mm. Bolozky 1999: 62–65). Tällaisissa testeissä kielenkäyttäjät tuottavat yleensä eniten helpoiten mieleen tulevia, läpinäkyvimpiä ja kielessä hallitsevimpiä rakenteita. Lisäksi kielenkäyttäjät suosivat tuottamistesteissä yleensä sellaisia rakenteita, joilla on mahdollisimman laajat semanttiset tulkintamahdollisuudet, vaikka tarkempi ja normatiivisesti soveliaampikin vaihtoehto olisi tarjolla. Toisinaan koehenkilöt tuottavat testitilanteessa myös yksilöllistä luovuutta heijastavia sanoja, jotka saattavat kuulua vain yhden ihmisen tai ihmisryhmän kielimuotoon. (Mts. 51.) Tuottamistestien tulokset voivatkin vaihdella eri koehenkilöiden välillä paljon. Mikäli tutkimuksen tuloksista tahdotaan tehdä esimerkiksi kokonaista kieltä tai kielimuotoa koskevia yleistyksiä, on tärkeää, että testauksessa käytetään riittävän suurta koehenkilöjoukkoa, jonka jäsenet edustavat erilaisia ihmisryhmiä (Bauer 2001: 57).

Toisenlaisia produktiivisuuden mittaamisessa käytettyjä kielenkäyttäjille tehtäviä testejä ovat leksikaaliset päätöksentekotehtävät (*Lexical Decision Task*), joissa koehenkilön tulee joko valita annetuista vaihtoehtoista mielestään paras tai arvioida käytetyn sanan kielenmukaisuutta tai esimerkiksi kuulumista omaan sanavarastoonsa. Yksi varhaisimmista leksikaalisia päätöksentekotehtäviä hyödyntävistä produktiivisuustutkimuksista on Mark Aronoffin ja Roger Schvaneveldtin (1978) englannin samanmerkityksisten *-iveness-* ja *-ixity-*johdinten produktiivisuutta käsittelevä artikkeli. Nykyaikaisissa päätöksentekokokeissa voidaan koehenkilön tietoisten ratkaisujen lisäksi tarkastella myös hänen tiedostamattomia reaktiotaan (esim. ERP-mittauksista produktiivisuuteen liittyvissä tutkimuksissa ks. esim. Havas–Rodríguez–Fornells–Clahsen 2012).

Keskeinen kysymys leksikaalisten tuottamis- ja päätöksentekokokeiden suunnittelussa on koeasetelmien ja koehenkilöille esitettävien kysymysten valinta (esim. Bolozky 1999: 51). Bolozky (mts. 51, 63) käsittelee erilaisten produktiivisuuden mittaamisessa käytettyjen menetelmien yhtäpitävyyttä tarkastelevassa tutkimuksessaan muun muassa koeasetelmaa, jossa koehenkilön tulee valita hänelle annetuista

vaihtoehtoista kontekstiin sopivin. Hänen mukaansa kielenkäyttäjä voi tällaisessa testissä saada eteensä sellaisiakin vaihtoehtoja, jotka eivät olisi tulleet hänen mieleensä tai joista hän ei välttämättä olisi ollut edes tietoinen, ellei niitä olisi osoitettu hänelle. Esimerkiksi siksi valittujen etukäteen odotuksenmukaisimpien (*obvious default*) vaihtoehtojen ja muiden vaihtoehtojen määrän välillä onkin tällaisessa testissä usein pienempi ero kuin tuottamistesteissä. Toisaalta tuottamistesteissä toisinaan esiintyvät marginaaliset, mahdollisesti vain yhden tai muutaman ihmisen idiolektiin kuuluvat vaihtoehdot jäävät yleensä pois asetelmissa, joissa koehenkilön tulee valita paras hänelle annetuista vaihtoehtoista. Tätä kautta testiasetelma voi ohjata joitakin koehenkilöitä ja siten myös testin tuloksia kohti etukäteen odotuksenmukaisimpana pidettyä rakennetta ja nostaa sen jopa suositummaksi kuin tuottamistestissä.

Bolozkyn (1999: 191) mukaan produktiivisuuden mittaamisessa käytettävät kielenkäyttäjien testausmenetelmät tarvitsevat edelleen kehittämistä erityisesti siinä, kuinka kielenkäyttäjiltä voitaisiin saada mahdollisimman luonnollisia ja spontaaneja vastauksia. Nähdäkseni tämä tarkoittaa esimerkiksi sitä, että testauksessa olisi suosittava tuottamistehtäviä päätöksentekotehtävien sijaan. Tuottamistesteissäkin päästäisiin todennäköisesti parhaisiin tuloksiin, jos koehenkilöiden huomio pystyttäisiin kiinnittämään johonkin muuhun asiaan kuin tietoiseen uusien sanojen luomiseen (mp.). Puutteistaan huolimatta myös kielenkäyttäjille tehtäviä päätöksenteko- ja tuottamistestejä on pidetty produktiivisuuden tutkimuksessa käyttökelpoisina varsinkin yhdessä muiden metodien kanssa käytettynä (Aronoff ja Schvaneveldt 1978: 112; Bolozky 1999: 191–194).

Yhdyssanarakenteiden tutkimuksessa testiä, jossa koehenkilön tulee valita paras annetuista vaihtoehtoista, voitaisiin käyttää ensisijaisesti silloin, kun vaihtoehtoisia rakenteita on kaksi tai korkeintaan muutamia. Tutkittavaksi sopiva kohde voisi olla esimerkiksi yhdyssanan alkuosan taipuminen (*isosisko : isosiskoa ~ isoasiskoa*, ks. myös 5.3.3). Todellisen kielenkäytön havainnointiin verrattuna testin etu olisi se, että sen avulla saataisiin helposti riittävästi materiaalia kielessä melko harvinaisen ilmiön tutkimiseen. Sen sijaan esimerkiksi yhdyssanan alkuosan eri muotojen produktiivisuuden mittaamiseen valintatesti ei sovi, sillä mahdollisia vaihtoehtoja on liikaa (ks. 5.3.2) ja marginaaliset vaihtoehdot jäisivät siksi väistämättä syrjään. Lisäksi on huomattava, että koska sama asia voidaan ilmaista monin eri sanoin ja koska esimerkiksi eri sanaluokkia edustavat sanat voivat esiintyä yhdyssanan alkuosana toimiessaan vain tietyissä muodoissa, jo yhdyssanan osina käytettävien sanojen valinta on osa kielenkäyttäjän tekemää rakenteen valintaa. Testi, jossa kielenkäyttäjän on valittava annetuista vaihtoehtoista paras, ohittaa tämän yhdyssanojen muodostukseen liittyvän olennaisen vaiheen.

Tuottamistehtävien testien käyttö yhdyssanarakenteiden tutkimuksessa on ongelmallista erityisesti siksi, että on vaikeaa rakentaa testiasetelmaa, jossa koehenkilöt loisivat spontaanisti nimenomaan uusia yhdyssanoja. On esimerkiksi vaikeaa keksiä sellaisia aukkolauseita, jotka olisivat täydennettävissä ainoastaan yhdyssanoilla. Jos taas koehenkilöitä pyydetään käyttämään aukkotestissä nimenomaan

uusia, itse keksittyjä yhdyssanoja, heidän huomionsa kiinnittyy väistämättä nimenomaan luovaan sanankäyttöön.

Toinen mahdollinen tapa rakentaa yhdyssanojen tuottamista mittaava testi-asetelma olisi mukaelma kielten opetuksessa käytettävistä sanojen yhdistämistehdävistä. Tehtävässä koehenkilölle annetaan joukko sanoja, joista hänen tulee valita yhteensopivat ja luoda niistä yhdyssana. Tällaisten tehtävien tarkoitus on yleensä harjoitella opiskeltavassa kielessä (äidinkieli tai vieras kieli) olevia vakiintuneita yhdyssanoja, mutta koehenkilöt voitaisiin ohjeistaa luomaan testissä myös uusia yhdyssanoja. Tämän testin ongelma on kuitenkin sama kuin aukkotehtävän: jotta koehenkilö saadaan tuottamaan uusia yhdyssanoja, hänen huomionsa on kiinnitettävä nimenomaan uusien sanojen tuottamiseen, jolloin sananmuodostus ei ole enää spontaania. Lisäksi muoto, jossa sanat annetaan koehenkilölle, voi ohjata vastauksia. Esimerkiksi sanojen esittely nominatiivissa voi ohjata koehenkilöitä tuottamaan enemmän nominatiivialkuisia yhdyssanoja, kuin he spontaanissa kielenkäyttötilanteessa tekisivät.

3.3.4 Tämä tutkimus:

Uudissanastossa havaittu rakenteen tyyppifrekvenssi

Tässä tutkimuksessa käytän metodia, jota ei ole tällaisenaan käytetty aiemmin produktiivisuuden tutkimuksessa. Uudissanastossa havaittua rakenteen tyyppifrekvenssiä⁴³ hyödyntävän metodin peruserä on yksinkertainen: Huolella valitussa, uutta sanastoa sisältävässä aineistossa eri rakenteiden frekvenssijakauma vastaa rakenteiden frekvenssijakaumaa kaikkien uusien sanojen joukossa. Tieto rakenteen suosiosta uudessa sanastossa puolestaan heijastaa rakenteen produktiivisuutta silloin, kun tarkasteltavana on produktiivisuuden kahdesta puolesta nimenomaan käytännöllinen produktiivisuus.

Metodi pohjautuu perinteiseen ajatukseen, jonka mukaan rakenteen frekvenssi kielessä osoittaa sen produktiivisuuden. Pelkästään kielessä esiintyviä rakenteita laskemalla ei kuitenkaan voida erottaa synkronista ja diakronista produktiivisuutta toisistaan, vaan vain tarkastella rakenteen yhteensä kaikkina aikoina tuottamia ilmauksia (ks. luku 3.1.2). Jos tarkasteltava aineisto kuitenkin rajataan niin, että se sisältää vain tiettyä aikana kieleen tullutta sanastoa, menetelmän avulla voidaan hyvinkin kuvata eri rakenteiden produktiivisuutta juuri tuon sanaston syntyhetkellä.

Edellä esitellyistä (3.1.1–3.1.3) aiemmassa tutkimuksessa käytetyistä metodeista tämän tutkimuksen metodi on teoreettisesti lähimpänä eri-ikäisten sanastojen

43 Rakenteen *tyyppifrekvenssillä* (*type frequency*) tarkoitetaan sitä, kuinka paljon tutkimusaineistossa on tiettyä rakennetta noudattavia eri ilmauksia. Sanastontutkimuksessa tyyppifrekvenssi on siis rakennetta noudattavien lekseemien määrä, ja rakenteen tyyppifrekvenssistä käytetäänkin toisinaan myös nimityksiä *leksikaalinen frekvenssi* ja *sanakirjafrekvenssi*. *Tapausfrekvenssillä* (*token frequency*) puolestaan tarkoitetaan sitä, kuinka usein tietty rakenne toistuu tekstissä. Sanastontutkimuksessa tapausfrekvenssi on siis rakennetta noudattavien saneiden määrä. Koska tämän tutkimuksen aineisto on tietokanta, jossa jokainen sana toistuu vain kerran, tutkimuksessa ei tarkastella tapausfrekvenssiä.

vertailuun perustuvaa tutkimusotetta. Suurin ero on metodien näkökulmassa. Eri-ikäisten sanastojen vertailun avulla voidaan tarkastella lähinnä tiettyjen rakenteiden produktiivisuuden diakronisia muutoksia. Uudessa sanastossa esiintyvien rakenteiden tyyppifrekvenssien tarkastelulla puolestaan etsitään vastausta ensisijaisesti kysymykseen siitä, mitkä keskenään vaihtoehtoisista rakenteista ovat kielenkäyttäjien suosiossa tällä hetkellä.

Esimerkiksi Plag (1999: 111–115) on korostanut, että luotettavien tulosten takaamiseksi produktiivisuuden tutkimuksessa käytettyjä kvantitatiivisia metodeja on täydennettävä kvalitatiivisella tarkastelulla. Uudessa sanastossa havaittuihin rakenteen tyyppifrekvensseihin perustuva metodi ei ole tässä suhteessa poikkeus, vaan siinä kvalitatiivista tarkastelua voidaan oikeastaan pitää analyysin toisena – ja välttämättömänä – askeleena. Rakennetta noudattavien uusien sanojen rakenneosia ja muodostustapoja on analysoitava erityisen tarkasti silloin, kun rakenne tuottaa vain vähän uusia muodosteita. Näin joukosta pystytään erottamaan sellaiset rakenteet, jotka eivät ole rakenteina sinänsä produktiivisia, vaan niitä noudattavien muodosteiden karttumisen johtuu jostakin muusta syystä. Muita syitä voivat olla esimerkiksi yksittäisiin sanoihin perustuvat analogiat sekä vain yhteen tai muutamaan sanaan liittyvät kiteymät, joissa rakenne on vakiintunut vain tiettyjen yksittäisten lekseemien yhteydessä esiintyväksi. Tällöin kyse on ennemmin noiden lekseemien kuin rakenteen produktiivisuudesta⁴⁴ (ks. myös luku 3.1.1).

Periaatteessa käytännön suosioon keskittyvältä tutkimukselta jäävät havaitsematta sellaiset rakenteet, joilla kyllä on kieliopilliset mahdollisuudet produktiivisuuteen mutta jotka eivät syystä tai toisesta tuota uutta. Esimerkiksi Bauerin (2001: 21) mukaan pelkät frekvenssit eivät siksi riitä produktiivisuuden tarkasteluun. Itse näen produktiivisuuden kuitenkin ilmiönä, josta voidaan erottaa kaksi puolta, kieliopillinen potentiaali ja käytännön suosio (ks. myös luku 3.1.2). Tässä tutkimuksessa huomio on kohdistettu rakenteen käytännön suosioon sananmuodostuksessa, sillä produktiivisuuden puolista ainoastaan käytännöllistä produktiivisuutta pystytään ylipäänsä tutkimaan empiirisesti autenttisen kieliaineiston avulla. Muita syitä näkökulman valinnalle ovat seuraavat: a) yhdyssanojen muodostus on suomen kielessä kieliopillisesti varsin vapaata ja muodostukseen liittyvät kieliopilliset rajoitukset tunnetaan jo hyvin (esim. ISK § 398–437), b) käytännön suosion tarkastelu tuo kieliopillisen potentiaalın tarkastelua paremmin esiin myös produktiivisuuteen liittyvät sosiaalis-tyylilliset ja prosessointia koskevat seikat ja c) rakenteen käytännön suosio kertoo kielen muutossuunnista sen kieliopillista potentiaalia enemmän.

Jos yhdessä tutkimuksessa tahdotaan tarkastella produktiivisuuden molempia puolia, uudissanastosta havaittu rakenteen tyyppifrekvenssi vaatii seurakseen muita metodeja. Kuten todettua, monet tutkijat suosittelevat produktiivisuuden tutkimuksessa ylipäänsä useiden metodien rinnakkaista käyttöä (mm. Bolozky 1999: 194; Plag 1999: 117; Bauer 2001: 158–161). Koska muiden aiemmin käytössä

44 Tietyn vakiomuotoisen lekseemin sisältävää muuttia voidaan toki myös kutsua rakenteeksi (esim. Tyysteri 2011c). Tässä tutkimuksessa näin ei kuitenkaan tehdä.

olleiden metodien ja niihin kuuluvien aineistotyyppien käyttöön yhdyssanarakenteiden tutkimuksessa liittyy kuitenkin ylikäymättömiä ongelmia (ks. luvut 3.3.1–3.3.3), olen päätenyt käyttämään tässä tutkimuksessa vain yhtä metodologiaa varsinaisessa produktiivisuuden mittaamisessa. Tässä tutkimuksessa käytetyn ja aiempien metodien yhtäpitävyyden tarkastelu jää siten esimerkiksi johtomorfologiaa käsitteleville tutkimuksille.

Uudessa sanastossa havaitun rakenteen tyyppifrekvenssin ja sitä tukevan kvalitatiivisen analyysin avulla voidaan etsiä vastausta ensisijaisesti kysymykseen siitä, onko jokin rakenne ylipäänsä produktiivinen kielenkäytössä eli muodostetaanko tiettyä rakennetta noudattavia yhdyssanoja käytännössä vai ei. Tyyppifrekvenssin avulla voidaan tehdä havaintoja myös produktiivisuuden asteesta, vaikka suhteellista frekvenssiä lukuun ottamatta yksin sen avulla ei voidakaan määrittää mitään tarkasti rakenteen produktiivisuutta kuvaavaa vertailulukua. Metodien avulla voidaan tarkastella myös eroja keskenään aidosti vaihtoehtoisten rakenteiden produktiivisuudessa. Vaikka tämä tutkimus on luonteeltaan synkroninen, vertaamalla tässä tutkimuksessa saatuja tuloksia suomen kielen yhdyssanoista aiemmin todettuihin seikkoihin, myös arvioita esimerkiksi jonkin rakennepiirteen yleistymisestä tai harvinaistumisesta voidaan esittää.

Jos käyttöön saadaan vertailukelpoiset aineistot eri aikakausilta tai kielimuodoista, metodien avulla voidaan tutkia myös produktiivisuuden diakronisia muutoksia tai eroja eri kielimuodoissa. Tällöin voidaan etsiä vastausta myös esimerkiksi Bauerin (2001: 155–158) produktiivisuudentutkimuksessa keskeisenä pitämään kysymykseen siitä, kuinka suuri osa tiettyä rakennetta noudattavista sanoista on uusia. Koska metodi hyödyntää suhteellisia frekvenssejä, vertailtavien aineistojen ei välttämättä tarvitse olla samankokoiset. Erikokoisten aineistojen mahdolliset vaikutukset tuloksiin on kuitenkin otettava huomioon: suurissa aineistoissa yleensä myös variaatio on suurempaa, ja lisäksi monet marginaaliset rakenteet tulevat ylipäänsä näkyviksi vasta hyvin suurissa aineistoissa. Lisäksi analyysia on aina syytä täydentää kvalitatiivisilla huomioilla. Suomessa diakroninen tutkimus on mahdollista viimeistään joidenkin vuosien tai vuosikymmenen kuluttua, kun vuonna 1984 kerätyt ja sitä myöhemmät sanatiedot sisältävä ja edelleen jatkuvasti kartutettava Nykysuomen sanastotietokanta eli NSST tulee riittävään ikään (NSST:sta tarkemmin luku 4.1).

Uudissanastossa havaittua rakenteen tyyppifrekvenssiä hyödyntävän metodien avulla saatujen tutkimustulosten luotettavuuden kannalta olennaista on nimenomaan uuden sanaston tunnistaminen. Käsitellen uusien yhdyssanojen tunnistamista tarkemmin seuraavassa luvussa, joka kuvaa tutkimuksen käytännön toteutusta.

4 Tutkimuksen käytännön toteutus

Tässä luvussa esittelen tutkimukseni aineiston ja työn käytännön toteutustavat sekä perustelen niiden valinnan. Kuten edellä on jo käynyt ilmi (luku 3.3), produktiivisuuden tutkimuksessa on käyty viime vuosikymmeninä vilkasta keskustelua siitä, millaiset aineistot ja metodit sopivat parhaiten aihepiiriin tutkimukseen. Tutkimusaineiston rajauksen ja tutkimuksen käytännön toteutuksen kuvaus ovat tietysti olennainen osa jokaista hyvää tutkimusta, mutta tässä työssä ne ovat erityisen tärkeitä, koska sekä käyttämäni aineisto että metodi ovat uudenlaisia. Tutkimuksen tärkeimmät käytännön menetelmät ovat tilastollinen kuvailu ja testaus. Myös niiden vuoksi tutkimusaineiston valinta ja rajausta sekä tutkimusmenetelmien valinta vaativat työssäni huolellisia perusteluja, sillä tilastollisen tutkimuksen hyödyllisyys ja käytännön relevanssi juontuvat ensisijaisesti siitä, että aineisto on laadukasta ja metodit tutkimusaineistoon soveltuvia (Luojola 2006: 17).

Tutkimusaineiston pohjana on Kotimaisten kielten tutkimuskeskuksen Nykysuomen sanastotietokanta (NSST), johon kerätään jatkuvasti uutta ja uudella tavalla käytettyä sanastoa lähinnä sanakirjatyon ja kielenhuollon tarpeisiin. Esittelen aineistolähteen tarkemmin tämän luvun ensimmäisessä alaluvussa (4.1). Käsittelem ensimmäisessä alaluvussa myös uudissanojen tunnistamiseen liittyviä kysymyksiä, sillä tässä tutkimuksessa käytetyn metodin ehdoton edellytys on, että tutkimusaineisto on nimenomaan uudissanastoa (ks. myös luku 3.3.4). Toisessa alaluvussa (4.2) kuvaan varsinaisen tutkimusaineistoni koostamiskriteerit. Aineistonrajausten tarkka esittely on olennaista erityisesti siksi, että tutkimustuloksia voidaan myöhemmin luotettavasti verrata toisenlaisiin aineistoihin. Kolmannessa alaluvussa (4.3) esittelen aineistoon koodatut muuttujat sekä annotointiprosessin. Tarkasteltavien muuttujien ja kategorioiden tarkka rajaaminen on edellytys luotettavien kvalitatiivisten tulosten saamiselle (McEnergy–Wilson 1996: 62–63; Luojola 2006: 10). Neljännessä alaluvussa (4.4) kuvaan aineiston analysoinnissa käytetyt tilastolliset menetelmät.

4.1 Aineistolähde: Nykysuomen sanastotietokanta

Tutkimukseni aineistolähde on Kotimaisten kielten tutkimuskeskuksen Nykysuomen sanastotietokanta (NSST), jonka pohjalta olen koostanut oman tutkimukseni tarpeisiin sopivan yhdyssanastotietokannan. NSST on osa Kotuksen nykysuomen sanakokoelmaa, ja sitä kartutetaan jatkuvasti. Vuosittain uusia poimintoja tehdään noin 10 000–15 000 (Grönros 1991: 33; Haarala–Nissinen 1994: 23). NSST sisältää hakusanan lisäksi tiedot sanan poimintapaikasta ja ajasta. Useimmista hakusanoista on lisäksi ilmoitettu vähintään yksi käyttökonteksti. Lisäksi joidenkin sanojen yhteyteen on merkitty hajanaisia tietoja sanan tyylilajista, alkuperästä sekä kuulumisesta jonkin erikoisalalan kieleen (esim. lääketiede, teletekniikka ja urheilu). Lisäksi mukana on tietoja siitä, kuuluuko sana johonkin poikkeukselliseen kieliopilliseen luokkaan (esim. yhdysverbit). NSST:aan sanastetaan sanoja pääasiassa kirjoitetusta kielestä, mutta poimintoja tehdään satunnaisesti myös puhutusta kielestä. Paljon käytettyjä lähteitä ovat esimerkiksi Helsingin Sanomat (HS), Aamulehti, Suomen kuvalehti ja Talouselämä.

NSST:n poimintakriteerien mukaan tietokantaan kerätään sellaisia sanoja (ja muita ilmauksia), jotka eivät ole aiemmin olleet yleiskielessä yleisessä käytössä tai joita on käytetty aiemmasta käytöstä poikkeavasti (Joki 2009). NSST:n tavoitteena on saada suomen kielen uudet sanat ja vanhojen sanojen uudet käytöt kerättyä talteen mahdollisimman laajasti (Haarala–Nissinen 1994: 23). Ensisijaisesti sanastoa kerätään sanakirjatyön ja kielenhuollon tarpeisiin (Eronen 2007: 21), mutta keräytystä aineistosta vain murto-osa päättyy lopullisiin sanakirjoihin. Koska uudissanat poimittaessa ei voida vielä tietää, mitkä sanat tulevat vakiintumaan kieleen niin, että ne on joskus syytä ottaa sanakirjaan, NSST sopii aineistolähteeksi myös vakiintumattoman sanaston tutkijalle.

NSST on ainutlaatuinen aineisto, jolla on varsinkin yhdyssanatutkimuksessa kiistattomia etuja suhteessa muihin produktiivisuuden mittaamisessa käytettyihin aineistoihin (aineisto ja metodit kietoutuvat produktiivisuuden tutkimuksessa vahvasti yhteen, ks. luku 3.3). Olen koonnut eri aineistotyyppien hyviä ja huonoja puolia tiivistetysti asetelmaan 1. Hyvät puolet on merkitty taulukossa plusmerkillä ja huonot puolet miinusmerkillä.

ASETELMA 1. Erilaisten aineistotyyppien soveltuvuus yhdyssanarakenteiden produktiivisuutta käsitteleviin tutkimuksiin (suomen kielestä saatavissa olevat aineistot)

	Sanakirjat	Tekstikorpuksset	NSST
Sisältää harvinaisia sanoja	- Vain vähän	+ Kyllä	+ Kyllä
Sisältää vakiintumattomia ⁴⁵ sanoja	- Ei	+ Kyllä	+ Kyllä
Sisältää läpinäkyviä sanoja	- Vain jos mukaan ottamiselle erityinen peruste	+ Kyllä	+ Kyllä
Kuua myös kieltä, joka ei ole norminmukaista	- Ei	+ Kyllä	+ Kyllä
Laajuus on riittävä yhdyssanojen morfologisen produktiivisuuden mittaamiseen	-/+ Ehkä ⁴⁶ (esim. NS, PS ja KS)	+ Kyllä	+ Kyllä
Voidaan käyttää ilman koneellista esikäsitteilyä	+ Kyllä	- Ei	+ Kyllä
Voidaan hallita ilman automaattisia analysointityökaluja	+ Kyllä	- Ei	+ Kyllä
Uudis-yhdyssanojen tunnistaminen on luotettavaa	- Ei (vertaamalla vanhempiin sanakirjoihin)	-/+ Ehkä (vertaamalla vanhempiin korpuksiin tai tulkitsemalla kertaesiintymät uudismuodosteiden määrän indikaattoriksi)	-/+ Ehkä (kielentutkijoiden intuitiolla tai vertaamalla vanhempaan aineistoon)

Kuten asetelmasta 1 nähdään, NSST:ssa yhdistyvät monet sanakirjojen ja tekstikorpuksien hyvät puolet: sanakirjojen tapaan NSST:aa voidaan käyttää ja hallita ilman automaattisia analysointityökaluja ja korpusten tapaan NSST:ssa on myös harvinaisia, vakiintumattomia ja läpinäkyvinä pidettyjä yhdyssanoja. NSST:n

45 Monet tutkijat erottavat varsinaiset kieleen jo jossain määrin vakiintuneet ns. varsinaiset uudismuodosteet tilapäismuodosteista ja kontekstisidonnaisista muodosteista. Yleensä sanan syntyhetkellä ei kuitenkaan voida ennustaa, ottaako kieliyhteisö sanan omakseen vai jääkö se päiväperhoksi. (Štekauer 2002.) Tässä tutkimuksessa kaikkia uusia yhdyssanoja kohdellaan samalla tavalla riippumatta siitä, ovatko ne jääneet yleiseen käyttöön vai eivät.

46 Ks. myös alaviite 39 sivulla 69.

toiseen pääkäyttötarkoitukseen eli kielenhuollon tarpeiksi tietokantaan poimitaan myös produktiivisuuden tutkijalle arvokasta norminvastaista sanastoa.

Kaikkien aineistotyyppien yhteinen ongelma on uudissanojen tunnistamisen epäluotettavuus, joka seuraa ennen muuta tarkkoja määritelmiä pakenevasta uudissanan eli neologismin käsitteestä (esim. Rey 1976). Kaikille nykyisille neologismin⁴⁷ määritelmille on yhteistä uutuuden ja tuoreuden liittäminen käsitteeseen, mutta uutuus ja tuoreus ymmärretään eri näkökulmissa eri tavalla.

Leksikografisesta näkökulmasta uusi on sana, joka puuttuu jostakin tietystä sanakirjasta tai valitusta sanakirjojen joukosta. Tämä näkökulma on ollut yksiselitteisyytensä vuoksi käytetyin uudissanatutkimuksissa (esim. Rondeau 1981: 137–138; Fischer 1998). Sanakirjoihin nojaava uutuuden määritelmä ei kuitenkaan sovi yhdyssanatutkimukseen, sillä varsinkin läpinäkyvinä pidetyt yhdyssanat jätetään usein tietoisesti sanakirjojen ulkopuolelle (ks. tarkemmin luku 3.3.1). Teresa M. Cabré (1999: 205) myös huomauttaa, että leksikografisen kriteerin käyttö on oikeastaan vain tapa siirtää vastuu uudissanojen valitsijalta sille, joka on koostanut vertailumateriaalin.

Tämän tutkimuksen aineistonvalinta nojaa sanan havaittavaan uutuuteen, joka on käsitteen ydin myös useissa neologismin määritelmissä. Kielenpuhuja tunnistaa uuden ilmauksen esimerkiksi kiinnittämällä huomiota sen epätavallisuuteen, harvinaisuuteen, yllättävyyteen ja tuoreuteen. (Esim. Rey 1995: 63–78; Cabré 1999: 205; Popova ym. 2005: 9; Eronen 2007: 22–24.) Ilmauksen uutuus kytkeytyy siis siihen, että kielenpuhuja kykenee havaitsemaan ilmauksen tunnusmerkkisyyden. Reyn (1995: 74) mukaan uutuuden ja tuoreuden tunnun pitäisi olla kieliyhteisön jakamaa, sillä muuten neologismina voidaan pitää myös sellaisia sanoja, jotka eivät ole uusia vaan vain joillekin kielenkäyttäjille ennestään vieraita.

Koko kieliyhteisön uudeksi tunnistamien sanojen tavoittaminen tutkimusaineistoksi ei ole yksinkertaista. Ratkaisuksi on esitetty uutena havaitsemisen testaamista suurella kielenpuhujien joukolla. Tällaiseen testaukseen liittyy kuitenkin useita käytännön ongelmia, joista hankalimpia ratkaistavia on se, että uutuuden tuntu katoaa monista sanoista nopeasti. (Cellard–Sommant 1979: 8; Popova ym. 2005: 9–11.) Nähdäkseni yksi parhaita keinoja tämän määritelmän mukaisten uudissanojen tavoittamiseen onkin se, että kielentutkijan koulutuksen saaneet ja sanaston tutkimukseen erikoistuneet ihmiset poimivat niitä intuiotensa nojalla elävästä kielestä. NSST on koostettu juuri tällä tavoin. Viime aikoina neologismien keräämiseen on kehitetty myös uudenlaisia automaattisia menetelmiä (tarkemmin esim. Paryzek 2008; Kerremans–Stegmayr–Schmid 2011), mutta suomen kieleen sopivaa työkalua ei ole toistaiseksi saatavilla.

47 Aiemmin neologismilla on tarkoitettu aivan muuta: sillä on viitattu esimerkiksi mieleltään järkyneiden ihmisten muiden kielenkäyttäjien kannalta merkityksettömiin tai väärässä merkityksessä käytettyihin sanoihin (Rey 1995: 63). Nykyäänkin neologismilla voidaan tarkoittaa myös eräissä afasian muodossa esiintyviä puheen häiriöitä, joissa ihminen tuottaa toisille ihmisille käsittämättömiä sanoja, jotka kuitenkin näyttävät olevan ymmärrettäviä hänelle itselleen (Laakkonen–Laakso 2001; Rohrer–Rossor–Warren 2009).

Toisinaan sanan voi paljastaa uudeksi sen merkitys tai motivaatio. Varsinkin yhdyssanojen kohdalla muodosteen uutuuden tunnistaminen edellyttää yleensä myös sen semantiikan tuntemista (Eronen 2007: 22–24). Jos sanalla viitataan vaikkapa uuteen keksintöön, itse sanakin on yleensä uusi (esim. *taulutietokone*). Tässä tutkimuksessa sanan uutuus määritellään sananmuodostusprosessin ajankohdan kautta, mutta denotatiivisen neologismikäsityksen mukaan ilmauksen uutuus voidaan jopa määritellä sen viittauskohteen kautta: denotatiivinen neologismi on uuteen tarkoitteeseen viittaava sana (Popova ym. 2005: 15).

Myös vanhoja sanoja otetaan toisinaan käyttöön uudessa merkityksessä (näin syntyneitä sanoja on kutsuttu semanttisiksi uudissanoiksi, esim. L. Hakulinen 1979: 383, 399; ks. myös Pitkänen 2008: 274). Tällöin uudella ja vanhalla merkityksellä on yleensä jonkinlainen semanttinen yhtymäkohta. Esimerkiksi vanhan selässä pidettävän kantovälineen *kontti*-nimityksen siirtymistä suuren rahtisäiliön nimitykseksi on perusteltu sillä, että sana muistuttaa äänteellisesti sen englanninkielistä *container*-vastinetta (Pulkkinen 1984: s. v. *kontti*; *konteineri*). Kuitenkin myös erilaisten kuljetusvälineiden semanttinen yhteys on selvä.

Periaatteessa myös yhdyssanat voivat esiintyä semanttisina uudissanoina, joissa vanha muodoste on siirtynyt uuteen merkitykseen. Uusiin merkityksiin ja merkityksen laajenemiseen liittyy usein kuvallisuutta (esim. 5). Tavallaan merkityksen laajenemisesta on kyse myös yhdyssanan uudenlaisessa käytössä termin osana (esim. 6).

- (5) Kaupungit ja kunnat haluavat mieluummin harmaita *betoniporsaita* kuin värikkäitä puutaloja. Ymmärrän tietysti, että puutalot kasvattavat tulipaloriskiä, mutta toisaalta ei ne betonitalotkaan ole tulipaloilta säästyneet. [TT <- Akaan Seutu, 2.1.7.2006, s. 8]
- (6) *Joustava eläkeikä* tarkoittaa sitä, että eläkkeelle pääsemisessä ei ole muita ehtoja kuin ikä. Jo 62 ikävuodesta valinta on jokaisen oma: voi lähteä eläkkeelle tai jatkaa työssään toistaiseksi – halutessaan jopa 68-vuotiaaksi asti. [TT <- HS 2.1.2005, s. A4]

Osiensa merkityksestä selvästi eriytyneet yhdyssanat eivät voi syntyä jokaisella käyttökerralla uudelleen, koska niiden vakiintunut merkitys on tunnettava entuudestaan, jotta niitä voi käyttää ja ymmärtää (esim. *kielikello*). Jos taas sanan merkitys on helposti pääteltävissä sen osista, ilmauksen voidaan ajatella syntyvän jopa jokaisella käyttökerralla uudestaan. Yhdyssanojen syntyhetken määrittely liittyykin kysymykseen siitä, kuuluvatko yhdyssanat ensisijaisesti sanastoon vai kielioppiin (ks. esim. A. Hakulinen – Karlsson 1979: 47; ks. myös luku 2.3.1).

Tämän tutkimuksen aineistoon on otettu kaikki sellaiset NSST:aan tarkasteltavana ajanjaksona (1.1.2000–1.9.2009) poimitut sanat, jotka ovat muodoltaan yhdyssanoja. Myös semanttisen sananmuodostuksen tietä syntyneet yhdyssanamuotoiset uudissanat kuuluvat siis tämän tutkimuksen aineistoon. Yleisempää kuin vanhan muodosteen siirtyminen uuteen merkitykseen on kuitenkin kokonaan uuden mutta vanhan yhdyssanan kanssa homonymisen uudismuodosteen luominen. Tämä on

seurausta siitä, että yhdyssanojen muodostaminen on hyvin produktiivinen prosessi ja lisäksi leksikaaliselta ja morfologiselta rakenteeltaan samankaltaisella sanalla voidaan viitata mihin tahansa sellaiseen tarkoitteeseen, jossa vain yhdyssanan osilla on jotakin tekemistä toistensa kanssa (ks. luku 2.3.1). Esimerkiksi sanalla *rullalauta* on aiemmin viitattu ainakin pöytämankeliin (Lönnrot 1958 [1866–1880] s. v. *rullalauta*), mutta nykyään sana on vakiintunut merkitykseen 'pyörillä varustettu pitkulainen levy, jolla liikutaan t. temppuillaan sen päällä seisten (vauhtia potkien) ja sitä painonsiirroilla ohjaten, skeittilauta, skeitti' (KS s. v. *rullalauta*).

Myös sama johdos voidaan luoda monta kertaa eri merkityksiin (esim. *eriö*-johdoksen eri syntytapauksista ja merkityksistä ks. Häkkinen 1990: 126). Yhdyssanojen erityispiirteiden vuoksi yhdyssanojen merkitysten muutosten ja sanan syntyhetken kuvaaminen on kuitenkin mutkikkaampaa kuin perussanojen tai johdosten, eikä uudessa merkityksessä käytettyjen vanhojen yhdyssanojen ja sattumalta vanhojen sanojen kanssa homonyymisten uusien yhdyssanojen erottaminen aina ole yksinkertaista. Esimerkiksi yhdyssana *ilmapallo* on vakiintunut merkitykseen 'leluna, koristeena tms. käytettävä pyöreämuotoinen ohutkuorinen ilma- t. kaasutäytteinen esine' (KS s. v. *ilmapallo*), mutta muodoltaan samanlaista sanaa voidaan käyttää esimerkiksi jalkapallo-ottelussa merkityksessä 'korkealle ilmaan potkaistu pallo' (ks. myös Räisänen 1986: 159). Nämä merkitykset vaikuttavat keskenään homonyymisiltä. Asia muuttuu mutkikkaammaksi, jos samaa ilmausta käytetään viittaamaan esimerkiksi leikkiin, jossa ilmapallo on tarkoitus pitää ilmassa niin pitkään kuin mahdollista. On vaikeaa ratkaista, onko tällöin kyse leikkivälineen nimen siirtymisestä leikin nimeksi vai viittaako nimitys leikin tavoitteeseen. Varsinkin leikittelevässä kielessä myös molempien motivaatioiden tunnistaminen saattaa rohkaista juuri *ilmapallo*-nimityksen valintaan leikin nimeksi.

Tässä uusia yhdyssanoja aineistonaan käyttävässä tutkimuksessa tukeudutaan kielentutkijan koulutuksen saaneiden kielenkäyttäjien intuitioon oman äidinkieltensä uusien ja uudella tavalla käytettyjen sanojen tunnistamisessa. Valtaosa NSST:aan poimituista yhdyssanoista on sekä morfologisia että semanttisia uudissanoja. Aakkosten alkupään (*j*-kirjaimen saakka) perusteella tekemäni arvion mukaan NSST:n yhdyssanoista noin 97–98 % on yleiskielen sanakirjaan verrattuna täysin uusia muodosteita sekä morfologisesti että semanttisesti. Vain noin 2–3 % sanoista esiintyy jossakin merkityksessä PS:ssa. Näistä sanoista noin kolmasosa on uudessa merkityksessä käytettyjä vakiintuneita yhdyssanoja (semanttisia uudissanoja), kolmasosa vakiintuneiden yhdyssanojen kanssa homonyymisiä uusia yhdyssanoja (morfologisia ja semanttisia uudissanoja) ja kolmasosa aineistoon syystä tai toisesta päätyneitä vakiintuneita yhdyssanoja. Vakiintuneet yhdyssanat ovat tämän tutkimuksen kannalta virhepoimintoja, mutta NSST:n päätarkoitusten eli sanakirja- ja kielenhuoltotyön kannalta niissä ei tietenkään ole mitään virheellistä.

Tämän tutkimuksen näkökulmasta virheellisiä poimintoja ei ole yritetty jälkikäteen rajata pois tutkimusaineistosta. Tällainen rajaus vaatisi ensinnäkin teoreettisen päätöksen siitä, miten jokaisella käyttökerralla uudelleen syntyvät ja pysyvinä leksikkoon kuuluvat yhdyssanat erotetaan toisistaan. Tätä päätöstä varten olisi

valittava, miten osien merkityksestä eriytymisen astetta mitataan ja mitkä kaikki merkitykset hyväksytään sanan vakiintuneeseen käyttöalaan kuuluviksi. Hyvätkään teoreettiset ratkaisut eivät jättäisi käytännössä muuta mahdollisuutta, kuin rajata aineisto sen perusteella, esiintyvätkö sanat aiemmissa sanakirjoissa ja onko niitä oman näkemykseni mukaan käytetty kontekstissaan samassa merkityksessä, joka sanakirjassa kuvataan (sanakirjojen ongelmista yhdyssanarakenteiden tutkimuksessa ks. myös luku 3.3.1). Sanakirjojen avulla tehtyä aineistonrajausta parempi vaihtoehto onkin tässä tutkimuksessa analyysin toisessa vaiheessa käytetty varsinkin harvinaisia rakenteita noudattavien sanojen kvalitatiivinen analysointi.

Tutkimuksen tulosten luotettavuutta arvioitaessa on pidettävä mielessä aineistolähteen ensisijaiset käyttötarkoitukset: sanakirjatyön ja kielenhuollon tarpeet. Ensisijaisten käyttötarkoitusten vuoksi aineiston koostamisessa käytetty menetelmä ei ole täysin satunnainen (ts. populaation jokaisella yksiköllä ei ole yhtä suuri mahdollisuus tulla valituksi aineistoon; lisää satunnaisotoksen kriteereistä esim. Luojola 2006: 24). Kielitieteessä epäsatunnaiset aineistot eli näytteet ovat kuitenkin usein ainoita järkevästi saatavissa olevia aineistoja, joten niiden käyttö on tavallista (Luojola 2006: 27), vaikka epäsatunnaisuuteen liittyikin vaara otoksen epäedustavuudesta (esim. Wang 1993). Puutteistaan huolimatta NSST on nähdäkseni tällä hetkellä paras aineistolähde suomen kielen yhdyssanarakenteiden produktiivisuuden tutkimukseen. Sen pohjalta koostetusta aineistosta saadut tulokset ovat tiettyin varauksin yleistettävissä kaikki suomen kirjoitetun yleiskielen uudet yhdyssanat käsittävään populaatioon (yleistyksiä tehdessä on syytä huomioida esimerkiksi ortografisten ja teoreettisten yhdyssanojen ero, jota on käsitelty luvussa 2.2.1).

4.2 Tutkimusaineiston rajaus

Olen saanut Kotuksesta käyttöni XML-muotoisena kaikki NSST:aan 1.1.2000–1.9.2009 tehdyt poiminnat (yhteensä 85 534 sähköistä sanalippua). NSST:ssa ei ole eroteltu yhdyssanoja perussanoista ja johdoksista, vaan olen koostanut tämän tutkimuksen tarkoituksiin sopivan Access-tietokannan NSST:n poiminnoista itse. Yhdestä yhdyssanasta voi olla NSST:ssä useampia poimintoja, mutta omassa tietokannassani jokainen sana esiintyy vain kerran. Lopullisessa tutkimusaineistossani on 28 091 yhdyssanaa.

Tutkimusaineistoni rajauksen pohjana olen käyttänyt luvuissa 2.1 esittelemääni (suomen kielen) prototyyppisen yhdyssanan kuvausta. Prototyypin kuvaus on tärkeä varsinkin yhdyssanan ja johdoksen välisessä rajanvedossa, jossa esimerkiksi ortografiasta ei ole apua. Yhdyssanan ja sanaliiton välisen rajanvedon olen puolestaan tehnyt ensisijaisesti sanan kirjoitusasun perusteella, sillä katson sen heijastavan kielenkäyttäjän omaa käsitystä ilmauksen yhdyssanuudesta (luku 2.2.1). Kirjoitusasuun perustuva rajaus on osin myös käytännön sanelema, sillä NSST:aan ei kerätä järjestelmällisesti ortografisia sanaliittoja, eikä sitä voida pitää niiden osalta

kattavana tai edes suuntaa antavana. Käytännössä aineistosta on jäänyt ortografian perusteella tehdyn rajauksen vuoksi pois joitakin sellaisia NSST:n poimintoja, jotka sopivat teoreettisen yhdyssanan määritelmään. Poisjääminen voi johtua joko teoreettisen yhdyssanan ja ortografisen yhdyssanan eroista tai siitä, että kirjoittaja ei ole tuntenut suomen kielen oikeinkirjoitussääntöjä. Esimerkiksi *skånen kieli* 'skånska eli Skånessa puhuttava kieli, joka ei [Skånen alueparlamentin mukaan] ole ruotsin murre' vertautuu teoreettiseen yhdyssanaan *suomen kieli*, joka suomen ortografian mukaan kuitenkin kirjoitetaan sanaliitoksi (ks. myös T. Kolehmainen 2014: 165–175). Puolestaan ilmauksessa *cross-over levy* teoreettista yhdyssanaa vastaava ja suomen ortografian mukaisesti yhteen kirjoitettava muodoste on kirjoitettu oikeinkirjoitussääntöjen vastaisesti sanaliitoksi ja se on siksi jäänyt pois tutkimusaineistosta.

Johdoksen ja yhdyssanan ero on useimmissa tapauksissa selvä. Prototyypisten tapausten jälkeen jäljelle jää kuitenkin vielä joukko sanoja, joiden kohdalla on käytettävä yksityiskohtaisempaa harkintaa ja tehtävä osin keinotekoisiaakin rajanvetoja. Jotkin elementit voivat toimia sananmuodostuksessa sekä johtimina että epäitsenäisinä leksikaalisina elementteinä (ks. esim. Rintala 1972: 25). Jotkin vierassananloput voivat esiintyä myös itsenäisinä sanoina (ks. luku 2.1.2). Mikäli (epäitsenäisenä) leksikaalisena elementtinäkin esiintyvä aines liittyy muodosteessa johtimen tapaan sanavartaloon (tai sen kaltaiseen ainekseen) ja noudattaa vokaalisointua, muodostetta ei ole otettu tutkimusaineistoon (esim. *sohvismi* 'joutilaisuutta korostava elämänfilosofia'). Jos taas elementti käyttäytyy muodosteessa lekseemin tavoin, se on otettu mukaan. Esimerkiksi sanassa *työholisti* [TT] loppuosa *holisti* esiintyy selvästi leksikaalisena elementtinä, sillä se ei ole mukautunut vokaalisointuun. Näin ollen koko muodoste on tulkittava yhdyssanaksi. Rajatapauksia ovat senkaltaiset ilmaukset kuin *simpnautti* 'avaruuteen lähetetty simpanssi', jossa on mahdollisesti pyritty johdosmaisuuteen, vaikkei alkuosa *simpa-* vastaakaan *simpanssi*-sanan kantaa. Tavoitteena on kuitenkin voinut olla myös yhdysosamaisen alkuosan muodostaminen. Myös tällaiset tapaukset on jätetty aineiston ulkopuolelle. (Tyysteri 2011b.)

Yhdyssanan sanaliitoista ja johdoksista erottamisen lisäksi uusien yhdyssanojen rajaamisessa on pohdittava ilmauksen ja sen osien kuulumista tiettyyn kieleen. Koska tämän tutkimuksen tarkoitus on tutkia nimenomaan suomenkielisisissä yhdyssanoissa esiintyvien rakenteiden produktiivisuutta, tutkimusaineistoon kuuluvat vain sellaiset yhdyssanat, jotka on muodostettu suomen kielen sananmuodostuksen periaatteita noudattaen ja joita voidaan pitää yhdyssanoina nimenomaan suomen kielen kannalta. Monet suomeen lainatut sanat ovat lainanantajakielessä yhdistesyntyisiä (esim. *arkkitehti*, *astronautti*, *biologia*) ja voivat hahmottua myös suomenkielisen mielessä kaksi- tai useampiosaisiksi⁴⁸. Osiksi hahmottumisesta huolimatta kokonaisuuksina suomeen lainattuja ilmauksia ei ole yleensä suomen

48 Sanakohtaiset erot ovat suuria. Iästä ja kielitaidosta riippumatta useimmat suomenkieliset vaikuttavat hahmottavan yksiosaiseksi esimerkiksi alkujaan yhdistesyntyisen mutta suomen kieleen kokonaisuutena lainatun sanan *internet*. Yhdyssanatulkintoja lisäävät sanan pituus, vokaalisoinnun rikkoutuminen sekä väliviiva sanassa. (Tyysteri 2011b.)

näkökulmasta pidetty yhdyssanoina. (Vesikansa 1989: 215; Klemettinen 2012: 12.) Sen sijaan käänöslainat (esim. *silmäkarkki* [TT] <- en *eye candy*⁴⁹) sekä sellaiset muodosteet, joissa lainasana yhdistyy kotoiseen sanaan (esim. *vadelma-chiboust* [TT], *wakeboard-lauta* [TT]) tai jotka on yhdistetty kahdesta tai useammasta lainasanasta vasta suomessa, ovat mukana aineistossa (ks. myös Häkkinen 1990: 162).

Rajaa kokonaisuutena lainatun ja lainatuista osista itse muodostetun yhdyssanan välillä voidaan etsiä esimerkiksi tarkastelemalla sanan ensiesiintymiä kielessä ja vertaamalla sen rakennetta muiden kielten vastaaviin tai lähes vastaaviin ilmauksiin. On kuitenkin otettava huomioon, että sama (yhdyss)ana voidaan muodostaa uudelleen monta kertaa (ks. luku 4.1). Esimerkiksi ilmauksella *bluesrock* on selvä vieras malli, mutta se on helposti muodostettavissa itsenäisesti myös suomen kieleen vakiintuneista aineksista. Tällaiset tapaukset olen ottanut aineistoon vain, jos niiden voidaan selvästi osoittaa syntyneen nimenomaan suomen kielessä tapahtuneen sananmuodostusprosessin kautta. Tästä voi olla osoituksena esimerkiksi osien mukautuminen suomen kieleen (esim. alkuosan loppu-*i* sanassa *gangsterijazz* [TT]). Olen jättänyt aineistoni ulkopuolelle myös sellaiset harvinaiset sanat, jotka voidaan kyllä jäljittää suoraan suomen kieleen luoduksi mutta jotka eivät noudata suomen sananmuodostuksen periaatteita tai sisällä suomen kieleen vakiintuneita lekseemejä⁵⁰ (esim. *finnhits*, *fifty-sixty*). Lisäksi olen jättänyt aineiston ulkopuolelle kemiallisten yhdisteiden ja muiden vastaavankaltaisten tarkoitteiden nimet riippumatta siitä, mihin kieleen ne on alkujaan luotu (esim. *darbepoietiini*, *parafenyleenidiamiini*, *polyeteenitereftalaatti*). Kemiallisten yhdisteiden nimenmuodostus ei noudata suomen sananmuodostuksen periaatteita, eikä aihetta tuntematon tavallinen kielenkäyttäjä todennäköisesti osaa analysoida niitä oikein osiinsa. Myös kielentutkijalta niiden tulkinta vaatii erityistietoja esimerkiksi kemian ja lääketieteen aloilta. Kemiallisten yhdisteiden nimien sisällyttäminen aineistoon olisi vääristänyt tuloksia erityisesti suomen kielen uusien yhdyssanojen leksikaalisten osien määrästä (ks. luku 5.1.1). Tämän vuoksi aineiston ulkopuolelle on jätetty myös sellaiset kemiallisten yhdisteiden nimet, joissa on kotoisiakin osia (esim. *dityppioksidi*, *rikkiheksafluoridi*).

Edellä mainittujen lisäksi tutkimusaineistoon eivät kuulu myöskään seuraavankaltaiset tapaukset:

- Ns. yksiosaiset yhdyssanat, joita käytetään elliptisissä ilmauksissa (esim. *myötä-* ja *vastoinkäymisissä*, *syntymäaika* ja *-paikka*) (tarkemmin esim. Penttilä 1963 [1957]: § 111, 221).

49 Suomessa on laajemmat mahdollisuudet yhdyssanan rakenteen valintaan kuin monissa lainantajakielissä. Periaatteessa esimerkiksi englannista mallinsa saaneissa käänöslainoissa voitaisiin käyttää vaikkapa genetiivialkua (esim. *silmänkarkki*). Käytännössä nominatiivialku on englannista mallinsa saaneissa käänöslainoissa kuitenkin lähes yksinomainen.

50 Yhtenä kriteerinä alkuaan vierasperäisen sanan kuulumisesta tiettyyn kieleen on käytetty sitä, käytetäänkö sanaa kyseisessä kielessä (Hiidenmaa 2003: 93–96). Väite suomenkielisydestä ei kuitenkaan tunnu käytännössä uskottavalta esimerkiksi sellaisten suomalaisten yritysnimien kohdalla, jotka rakentuvat täysin tunnistettavista englannin kielen sanoista (Sjöblom 2006: 111).

- Sellaiset esimerkiksi vaihteluvälejä, vaihtoehtoja ja alku- ja päätepisteitä ilmaisevat sanayhtymät, joita yhdistää yhdys-, ajatus- tai vinoviiva mutta joita ei yleensä pidetä sen enempää yhdyssanoina kuin sanaliittonakaan (esim. *kahteen-kolmeen viikkoon*, *Turun–Tampereen tie*, *joutuvat/pääsevät naimisiin* ks. ISK § 401; joissakin yhdyssanatutkimuksissa myös tällaiset tapaukset on sisällytetty aineistoon, esim. Lantee 2010).
- Deskriptiiviset äännejonot, joista ei aina voida yksiselitteisesti erottaa merkityksellisiä leksikaalisia osia, vaikka ne saattavat monien kielenkäyttäjien mielessä hahmottua kaksiosaisiksi (esim. *hohhoijaa*, *hokkuspokkustella* ja *tingelitangeli*). Sen sijaan mukana aineistossa ovat sellaiset muodosteet, joissa deskriptiivinen ilmaus tai interjektio yhdistyy selvän merkityksen omaavaan leksikaaliseen osaan (esim. *örbäturisti* [TT], *hurraa-vuodet* [TT]).
- Yhdysnimet ja nimistä johdetut yhdyssanat (esim. *Kultaturbo* [Suosikki-lehden erikoisnumeron nimi], *halla-aholaisuus*, *eppunormalisoitua*). Sen sijaan sellaiset yhdyssanat, joiden osista jokin on nimi mutta jotka eivät kokonaisuutena ole erisnimiä, on otettu mukaan aineistoon (esim. *Subutex-äiti* [TT], *väyrysilmiö* [TT]).

Lisäksi aineistoon valinnan kriteerinä on sanan käyttö painetun median toimitetussa tekstissä, sillä se sopii kielimuodoista parhaiten kirjoitettua yleiskieltä käsittelevän tutkimuksen aineistoksi. Aineiston ulkopuolelle olen siis jättänyt esimerkiksi mainoksista, kauno- ja tietokirjallisuudesta, internetistä ja puheesta NSST:aan poimitut sanat sekä sellaiset sanat, joiden yhteyteen ei ole kirjattu tietoa poimintapaikasta. Lisäksi olen poistanut sanat, joiden yhteyteen ei ole merkitty yhtään käyttök kontekstia. Näiden sanojen merkitystä on usein mahdoton tulkita, ja toisinaan myös rakenteen tulkinta jää epävarmaksi (onko esim. sanan *autonavigointi* [TT] alkuosa kulkuneuvon viittaavaa nominatiivimuotoinen substantiivi vai automaattisuuden viittaava prefiksinkaltainen elementti; ks. myös esimerkit 7 ja 8 sivulla 92).

Vaikka suomen kielen eri varieteeteissa ovat käytössä samat sananmuodostuskeinot, niiden produktiivisuudessa voi olla eroja kielimuotojen välillä (Pitkänen 2008: 127). Kirjoitettua yleiskieltä ja vielä spesifimmin lehtikieltä sisältävästä aineistosta saatuja tuloksia ei voida yleistää koskemaan muita kielimuotoja (esim. puhekieli, murteet, slangi, erikoiskielet). Tämän tutkimuksen tulosten ja suomen muista kielimuodoista ennestään tiedettyjen seikkojen perusteella voidaan kuitenkin esittää muiden suomen kielen varieteettien yhdyssanarakenteiden produktiivisuutta koskevia arveluita, joita voidaan myöhemmässä tutkimuksessa käyttää testattavina hypoteeseina.

Käytännössä samankin sanomalehden eri osastojen kielen on todettu poikkeavan toisistaan tilastollisesti (ks. esim. Niemikorpi 1990: 25–39, 1991). Tämän tutkimuksen tulokset ovat siksi täysin yleistettävissä vain lehtikieleen kokonaisuutena. Erityyppisten lehtien ja vaikkapa sanomalehden eri osastojen suhteen tuloksia voidaan pitää vain suuntaa-antavina, sillä niiden erityispiirteet jäävät aineiston

kokonaisanalyysissä pimentoon. Esimerkiksi joidenkin rakenteiden mukanaan kantama tyyliarvo (esim. useiden yhdysverbien leikillisuus) saattaa tehdä niistä vain joihinkin tekstilajeihin tyypillisesti kuuluvia.

Koska tämän tutkimuksen näkökulma on synkroninen ja koska tutkimuskohteenä on erilaisten yhdyssanarakenteiden suosio uudessa sanastossa, olen ottanut NSST:sta aineistooni vain sellaiset muodosteet, jotka tavallinen kielenkäyttäjä voi tunnistaa yhdyssanoiksi ilman tietoa esimerkiksi ilmauksen leksikaalisten osien historiasta. Valinta on kielen rakenteiden tämänhetkistä produktiivisuutta tarkastelevassa työssä luonnollinen, sillä se vastaa parhaiten sitä näkökulmaa, josta tavallinen kielenkäyttäjä kieltä katsoo. Esimerkiksi sanassa *outous* esiintyvän *-Us*-johtimen on arveltu olevan lähtöisin sanasta *vuosi* (Penttilä 1963 [1957]: § 219). Nykykielessä *-Us* käyttäytyy kuitenkin johtimen tapaan, eikä nykykielenkäyttäjä tunnista yhteyttä *vuosi*-sanaan. Näin ollen esimerkiksi sellaiset sanat kuin *asustus* ja *baittaus* '[grafittisanastossa] matkiminen, kopiointi' eivät kuulu tämän tutkimuksen aineistoon. Myös yhdyssanojen leksikaalisten osien määrää käsittelevässä luvussa (luku 5.1.1) olen noudattanut periaatetta, jonka mukaan tavallisen kielenkäyttäjän on tunnistettava yhdyssanan osat erillisiksi lekseemeiksi. Tämän periaatteen mukaan esimerkiksi *maailma* on laskettu yhdyssanan osana yhdeksi leksikaaliseksi elementiksi (ks. myös Tyysteri 2011b).

Tiivistäen voidaan sanoa, että olen valinnut NSST:sta tutkimusaineistooni vain sellaiset sanat, jotka sisältävät vähintään kaksi NSST:ssa yhteen kirjoitettua ja synkronisesta näkökulmasta leksikaaliseksi tunnistettavaa osaa. Lisäksi osista vähintään yhden on oltava suomenkielinen appellatiivi. Yhdyssanuuden kriteerinä on aineistonrajauksessa käytetty ainoastaan yhdyssanan rakennetta, ei sanan muodostustapaa. Näin ollen aineisto sisältää sekä yhdistämällä että johtamalla muodostettuja uusia yhdyssanoja. Käsitelen yhdyssanan muodostustavan ja rakenteen suhteita tarkemmin luvussa 6.1 (muodostustavoista ks. lisäksi luku 2.2.3). Aineisto on rajattu käsittämään ainoastaan toimitetussa lehtikielessä käytettyä sanastoa. Kaiken kaikkiaan tutkimusaineisto on varsin kattava otos 2000-luvun ensimmäisellä vuosikymmenellä suomen kirjoitettuun yleiskieleen syntyneitä yhdyssanas-toja ja tutkimuksen tulokset ovat nähdäkseni yleistettävissä koko populaatioon. Tutkimusaineiston rajausta arvioitaessa on kuitenkin syytä pitää mielessä, että se on laadittu nimenomaan käytännön tarpeisiin eikä sen ole tarkoitus olla esimerkiksi teoreettisten yhdyssanojen luokkaa täydellisesti rajaava klassisen kategorian määritelmä.

4.3 Aineiston annotointi ja hallinta

Annotointivaiheessa olen käsitellyt aineistoa pääasiassa Microsoft Access -tietokantana. Tilastollista kuvailua ja testausta varten olen muuntanut aineiston tekstitiedostoksi (.txt) ja vienyt RStudio-ohjelmaan (RStudio 2009–2013; tilastomenetelmistä

ks. seuraava luku 4.4). Aineisto on muunnettavissa helposti myös muihin tiedostomuotoihin (esim. XML- tai SQL-muotoihin).

Koska automaattiset annotointimenetelmät eivät ole suomen kielen yhdyssanojen osalta vielä toistaiseksi riittävän luotettavia (ks. luvut 3.3.2 ja 4.1), olen tehnyt valtaosan tutkimusaineiston annotoinnista käsin. Täysin koneellisesti Access-ohjelman hakutoimintojen avulla on tehty ainoastaan yhdyssanan ja sen osien merkimmäisiin laskemiseen. Lisäksi olen käyttänyt Accessin haku- ja suodatustoimintoja annotoinnin apuna sellaisissa tapauksissa, joissa yhdyssanojen rakenteessa on täydellisiä riippuvuussuhteita. Esimerkiksi *-in-* ja *-mA-*loppuisella verbimuodolla alkavien yhdyssanojen alkuosa on voitu koodata sanaluokaltaan verbiksi täysin mekaanisesti. Accessin lisäksi olen käyttänyt aineiston annotoinnin apuna Omorfi-ohjelmaa (Omorfi 2012), jonka avulla olen tehnyt kaksiosaisen nominatiivialkuisten yhdyssanojen jaon osiin. Aineiston annotointivaiheessa käytettävissä ollut Omorfi-versio ei osoittautunut riittävän luotettavaksi muulla kuin nominatiivimuodolla alkavien yhdyssanojen segmentoinnissa. Myös nominatiivialkuisten sanojen jaot on tarkastettu käsin. Omorfi 2012 ei osoittautunut käyttökelpoiseksi myöskään lemmatun yhdyssana-aineiston sanaluokan määrittämisessä.⁵¹

Tutkimustietokanta on jaettu kahteen osaan, joista toinen sisältää tiedot hakusanojen rakenteesta (varsinainen aineisto) ja toinen hakusanojen alkuperäisistä lähteistä ja käyttökonteksteista (alitetokanta). Tietokannan avaimena toimii hakusana. Haut on mahdollista kohdistaa myös hakusanan osiin, jotka on koodattu omiin sarakkeisiinsa. Asetelmassa 2 on esitelty tietokantaan koodatuista muuttujista ne, jotka ovat tutkimuksessa varsinaisen analyysin kohteena. Asetelmassa laittimmaisena oikealla olevassa sarakkeessa on listattu kaikki ne analyysiluvut, joissa kyseistä muuttujaa käsitellään. Muuttujan mahdolliset arvot esitellään tarkemmin ko. luvuissa. Asetelmassa esiteltyjen muuttujien lisäksi aineistotietokannassani on sarakkeita, jotka sisältävät esimerkiksi ala-aineistojen suodattamiseen tarvittavia tietoja sekä erilaisia huomautuksia.

Tutkimuksessa tarkasteltavaksi valitut muuttujat määrittävät sen, mitä tutkuskohde voidaan nostaa esiin. Myös käytettävä luokitus ja luokkien rajojen määrittely osoittavat osaltaan, mitä tutkimuksessa pidetään tärkeänä. Tässä tutkimuksessa tarkasteltaviksi on valittu sellaisia muuttujia, jotka ovat aiemman tutkimuksen valossa relevantteja yhdyssanojen rakenteen tutkimuksessa (ks. luku 2).

51 Erityisesti uudissanaston analysointi on automaattisilla työkaluilla epäluotettavaa. Sen lisäksi, että aineiston sanat ovat uusia yhdyssanoina, myös niiden osista monet ovat suomen kielelle uusia. Toinen automaattisten työkalujen käytön ongelma on leksikaalistumien huomioiminen (esim. *puhe-
linvanhemp* [TT] ei ole adjektiivi).

ASETELMA 2. Tutkimuksessa tarkasteltavat muuttajat ja niiden mahdolliset arvot

Muuttujan nimi	Muuttujan tyyppi	Muuttujan arvot	Käsittelyluvut
Hakusana (tietokannan avain)	kategorinen	Esimerkkejä mahdollisista arvoista: <i>etualamusiikki, osaamiskeskittymä, saappaanheittäjä, tasa-arvoajokortti</i>	Kaikki
Hakusanan leksikaaliset osat (osat 1–6 kukin omassa sarakkeessaan; 1. osa myös lemmattuna)	kategorinen	Esimerkkejä mahdollisista arvoista: <i>etu, ala, musiikki</i> <i>osaamis</i> [lemmattuna <i>osaaminen</i>], <i>keskittymä</i> <i>saappaan</i> [lemmattuna <i>saapas</i>], heittäjä <i>tasa, arvo, ajo, kortti</i>	Kaikki, kuitenkin erityisesti 6.2
Leksikaalisten osien määrä	numeerinen/ kategorinen ⁵²	2–6	5.1.1, 7.1
Koko muodosteen merkkimäärä	numeerinen	5–41	5.1.2, 7.1, 7.2
Ensimmäisen leksikaalisen osan merkkimäärä	numeerinen	1–27	5.1.2, 7.1, 7.2
Viimeisen leksikaalisen osan merkkimäärä	numeerinen	2–19	5.1.2, 7.1, 7.2
Koko muodosteen sanaluokka	kategorinen	adjektiivi, adjektiivi-substantiivi, adverbi, partisiippi, pronomini, substantiivi, verbi	5.2.2, 6.1.1, 6.2, 7.1.2, 7.2
Alkuosan sanaluokka	kategorinen	adjektiivi, adjektiivi-substantiivi, muu, numeraali, pronomini, substantiivi, sanaluokaton, verbi	5.2.4, 6.2, 7.2
Alkuosan muoto (karkea luokittelu)	kategorinen	epäitsenäinen elementti, genetiivi, muu sijamuoto kuin nominatiivi tai genetiivi, nominatiivi, taipumaton, yhdysosamuoto	5.3.1, 6.1.2, 6.2, 7.2
Alkuosan muoto (hienojakoinen luokittelu)	kategorinen	epäitsenäinen elementti (3 alakategoriaa) genetiivi (4 alakategoriaa) muu sijamuoto (9 alakategoriaa) nominatiivi (5 alakategoriaa) taipumaton (2 alakategoriaa) yhdysosamuoto (9 alakategoriaa)	5.3.2, 6.1.2, 6.2, 7.2
Osien kongruenssi	kategorinen (binäärinen)	kyllä, ei	5.3.3
Muodostustapa	kategorinen (binäärinen)	yhdistetty, johdettu	6.1

52 Muuttuja on periaatteessa numeerinen, mutta vähien mahdollisten arvojen vuoksi muuttujaa on osassa tilastollisia testejä kohdeltu kategorisena muuttujana, jolloin muuttujan arvot ovat 2-osainen, 3-osainen jne.

Koko tutkimusaineistoon kuuluu 28 091 yhdyssanaa. Analyysia varten olen erottanut tästä joukosta kulloisenkin tarkastelukohteen kannalta tarkoituksenmukaisimman osa-aineiston (ks. asetelma 3). Osa-aineistot on numeroitu, ja analyysissä niihin viitataan niiden numerolla. Osa-aineistoa 1 olen käyttänyt yhdyssanojen leksikaalisten osien määrän ja sanaluokan tarkasteluun, ja se sisältää kaikki tutkimusaineiston sanat lukuun ottamatta useammassa kuin yhdessä eri kirjoitusasussa esiintyviä ilmauksia, joista on mukana vain yksi kirjoitusasu. Osa-aineistoa 2 olen käyttänyt yhdyssanan merkkimäärän ja eri muuttujien keskinäisten suhteiden tarkasteluun. Aineistosta on jätetty pois välimerkkejä tai välilyöntejä sisältävät tapaukset, sillä välimerkit voisivat vääristää merkkimäärää koskevia tuloksia. Tutkimuksessa laajimmin käytetty osa-aineisto on osa-aineisto 3. Tähän aineistoon kuuluu vain kaksi leksikaalista osaa sisältäviä yhdyssanoja, ja siitä on jätetty pois esimerkiksi lyhenteellä tai sitaattilainalla alkavat ilmaukset. Tutkimuksessa käytetyistä muuttujista alkuosan muoto ja alkuosan sanaluokka on koodattu vain tähän osa-aineistoon, sillä aina ei ole selvää, mikä kolme tai useampia leksikaalisia osia sisältävän yhdyssanan osista on katsottava sen ensimmäiseksi yhdysosaksi (esim. *tekojäärata < tekojää + rata ~ teko + jäärata*, esimerkki Vesikansa 1989: 214).

ASETELMA 3. Tutkimusaineistosta erotetut osa-aineistot

Numero	Osa-aineiston kuvaus	Luvut, joissa käytetty
1	Aineiston kaikki sanat lukuun ottamatta niitä, jotka esiintyvät aineistossa kahdessa tai useammassa eri kirjoitusasussa (sanaluokka ja leksikaalisten elementtien määrä on kaikissa kirjoitusasuissa sama). Tällaisista sanoista on otettu mukaan vain yksi. Aineiston $n = 28\ 013$.	5.1.1, 5.2.2
2	Aineiston kaikki (ml. kahdessa tai useammassa kirjoitusasussa esiintyvät) sanat lukuun ottamatta niitä, jotka sisältävät lyhenteen tai välilyöntejä tai välimerkkejä (välilyöntien vaikutuksesta yhdyssanan tulkintaan Inhoff–Radach–Heller 2000; yhdysviivojen vaikutuksesta Bertram–Hyönä 2013). Aineiston $n = 25\ 903$.	5.1.2, 7.1
3	Aineiston kaikki kaksi leksikaalista osaa sisältävät sanat lukuun ottamatta seuraavia: - sanat, jotka esiintyvät useammassa kuin yhdessä kirjoitusasussa (mukaan otettu vain yksi), - sanat, joiden alkuosa on lyhenne ja - sanat, joiden alkuosa on edusosasta viivalla erotettu sitaattilainana. Aineiston $n = 22\ 144$.	5.2.4, 5.3.1, 5.3.2, 6.2, 7.2 (lisäksi luvussa 6.1 on käytetty satunnaisotoksia tästä osa-aineistosta)

Kvantitatiivisen tutkimuksen taustalle tarvitaan aina paljon kvalitatiivista työtä – aineiston annotointiperusteita koskevia päätöksiä ja varsinaista annotointia, jossa jokaista tapausa on pohdittava tarkoin ennen sen luokittamista. Kvalitatiivisessa tutkimuksessa yksittäisiä tapauksia voidaan pohtia pitkästikin eikä lopullista

luokituspäättöstä tarvitse välttämättä edes tehdä. Jos kategorioiden väliin jääviä rajatapauksia on paljon, käytettyä luokitusta voidaan muuttaa. Yleensä luokittelussa voidaan valita joko epävarma tarkka tulos tai varma epätarkempi tulos. Valmiiden kategorioiden välisten erojen ja rajatapausten lisäksi annotoinnissa nousee esiin myös kategorioiden sisäinen vaihtelu. Tässä tutkimuksessa aineiston luokittelussa on mahdollisuuksien mukaan käytetty aiemmassa tutkimuksessa ja kielitieteessä perinteisesti käytettyjä kategorioita, jotka helpottavat eri tutkimusten tulosten vertailua. Aineistolle oikeutta tekevä ja totunnaisten kategorioiden sisäisen vaihtelun riittävästi esiin tuova analyysi on kuitenkin vaatinut myös uusien tai totunnaisesta poikkeavien kategorioiden käyttöä.

Aineiston annotoinnissa tärkeimpänä luokitteluperusteena on käytetty yhdyssanan ja sen osien muotoa. Muodon lisäksi analyysi vaatii usein myös yhdyssanan ja sen osien merkityksen ja (mahdollisten) syntaktisten tehtävien huomioon ottamista. Esimerkiksi *elää*-verbin *-in*-loppuisella yhdysosamuodolla alkavan *elinkaari*-sanan [TT] 'tuotteen vaiheet markkinoilla kokonaisuutena' (KS s. v. *elinkaari*) ja nominatiivimuotoisella *elin*-substantiivilla alkavan *elinjärjestelmä*-sanan [TT] '(esim. sisä)elinten muodostama järjestelmä' keskenään homogرافisten alkuosien sanaluokan ja morfologisen muodon määrittely vaatii muodosteiden merkityksen tuntemista. Toinen kuvaava esimerkki merkityksen huomioon otamisen tärkeydestä on alkuosa *euro-*, joka voi koko muodosteen merkityksestä riippuen olla joko rahayksikköön viittaava nomini (esim. 7) tai Eurooppaan, eurooppalaiseen yhteistyöhön tai EU:iin viittaava sanaluokaltaan epäspesifi prefiksinkaltainen alkuosa (esim. 8).

- (7) Yleisöpuhelimet *eurokelpoisiksi* tammikuussa [otsikko] [TT <- HS 3.1.2002, s. B1]
- (8) *Euroarmeijan* jäsenmaiden kannattaisi Hägglundin mielestä myös erikoistua omiin vahvuuksiinsa. – – Suomalaisten vahvuuksiksi Hägglund laskee rauhanturvaamisen ja pioneeroitoinnin. [TT <- HS 8.9.2002, D 4]

Aineiston sanojen merkityksen selvittämisessä olen tukeutunut ensisijaisesti NSST:aan poimittuihin käyttökonteksteihin. Niiden lisäksi korvaamaton apu sanojen merkitysten selvittämisessä on ollut internet. Muita käytettyjä lähteitä ovat erilaiset sanakirjat (esim. KS, US, UD, Wikipedia FI sekä sen erikieliset rinnakkaissivustot). Aineiston sanojen merkityksiä ei ole tutkimuksessa selitetty järjestelmällisesti. Olen kuitenkin pyrkinyt antamaan merkityksenselityksen tai käyttökontekstin selailaisten sanojen yhteydessä, joiden merkitys on jollakin tavalla odotuksenmukaisesta poikkeava tai joiden olen arvellut herättävän lukijoissa kysymyksiä.

4.4 Tilastolliset menetelmät

Tutkimuksen analyysiosan tärkeimmät käytännön menetelmät ovat tilastollinen kuvailu ja testaus. Keskeinen syy tilastollisen näkökulman valitsemiselle on se, että haluan tässä tutkimuksessa esittää empiiriseen aineistoon perustuvaa tietoa eri yhdyssanarakenteiden produktiivisuudesta. Empiiriset ja varsinkin tilastolliset näkökulmat ovat huonosti edustettuina aiemmassa suomen kielen yhdyssanoja käsittelevässä tutkimuksessa.

Tilastollinen kuvailu ja testaus tukevat toisiaan. Tilastolliseen kuvailuun kuuluu aineiston ja sen muuttujien jakauman esittely numeerisesti ja graafisesti. Erilaisista graafisista kuvioista ovat tässä tutkimuksessa käytössä histogrammit, pylväsdiagrammit, laatikkojanakuviot ja luokittavat puukuvaimet⁵³. Aiheeseen perehtynyt tutkija näkee ryhmien erojen merkitsevyyden usein jo pelkästään vertailemalla jakaumia silmämääräisesti. Tilastollisen testauksen avulla asiantuntijan arviot voidaan osoittaa tosiksi myös tilastollisen päättelyn kautta: tilastollinen testaus auttaa päättelämään, onko havaittu ero seurausta jostakin siihen vaikuttavasta tekijästä vai voiko tulos olla sattumanvarainen. Tilastolliseen testaukseen on aihetta vain, jos tilastollisessa kuvailussa havaitulla erolla on käytännön merkitystä tutkimuksen aiheen kannalta (esimerkiksi sillä, onko yhdyssanan toinen kirjain tilastollisesti merkitsevästi useammin *e* kuin *a* ei ole käytännön merkitystä ainakaan tämän tutkimuksen aiheen kannalta). Tilastollinen testaus ei olekaan koskaan arvo sinänsä, vaan se toimii aiheen substanssiasiantuntemuksen tukena. (Wang 1993; ks. myös Luojola 2006: 90.)

Varsinkin suuren aineiston tilastolliseen testaukseen sisältyy riski siitä, että testit osoittavat tilastollisesti merkitseväksi sellaisiakin eroja, jotka ovat sattuman aikaansaamia (Wang 1993: 9). Tässä tutkimuksessa suuren aineiston aiheuttama harha on vältetty käyttämällä tilastollisessa testauksessa koko aineistosta otettuja pienempiä satunnaisotoksia⁵⁴. Koko tutkimuksen kannalta suuri aineisto on kuitenkin tarpeen, siellä suomen kielen yhdyssanoissa esiintyy monia harvinaisia rakenteita ja niiden yhdistelmiä, joita edustavat sanat putoavat helposti kokonaan pois satunnaisotoksista. Suurellakaan aineistolla ei voida osoittaa, että jotakin

53 Laatikkojanakuvioita käytetään luvussa 7.1 ja niiden tulkitsemisesta kerrotaan tarkemmin ensimmäisen laatikkojanakuvion yhteydessä luvussa 7.1.1 sivulla 176 kuvio 7. Luokittavia puita käytetään luvussa 7.2, mutta puukuvaimien tulkitsemiseen opastetaan jo tässä luvussa.

54 Toinen vaihtoehto olisi ollut valita tilastolliseen testaukseen alhaisempi merkitsevyytaso. Koska tutkimuksessa vertaillaan myös aineiston sisällä olevia pieniä ryhmiä, koko tutkimuksessa on johdonmukaisuuden vuoksi käytetty samaa merkitsevyytaso (α = 0,05). Testauksessa käytettyjen otosten maksimikoko on 1000 tapausta. Eri ryhmien vertailussa otoskoot on valittu pienimpien ryhmien koon mukaan. Otoskoko on mainittu kunkin testin yhteydessä. Aineistosta otetuista pienemmistä satunnaisotoksista raportoidut tilastolliset suureet ovat keskiarvoja kymmenellä eri otoksella tehdyistä testeistä.

A. Hakulinen, Karlsson ja Vilkuna (1996 [1980]: 102–104) esittelevät tarvittavan aineiston minimikoon määrittämiseen sopivan menetelmän. Heidän tutkimuksensa tarkoitus on kielitieteellisten otantojen metodiikan tarkentaminen. Tässä tutkimuksessa ei ole olennaista, että tilastollinen testaus tehdään juuri minimikokoisilla otoksilla.

ilmiötä ei olisi olemassa lainkaan, mutta suurella aineistolla on silti pientä aineistoa paremmat mahdollisuudet tavoittaa marginaalisetkin tapaukset⁵⁵. Toisinaan poikkeuksellisilla havainnoilla voi myös olla tavanomaisia tapauksia suurempi tieteellinen merkitys. Suuren aineiston toinen hyöty on, että sen avulla melko harvinaisistakin tapauksista saadaan riittäväkokoisia ryhmiä tilastolliseen testaukseen. Tämä mahdollistaa matalafrekvenssistenkin piirteiden keskinäisten vaikutussuhteiden tarkastelun.

Tilastollisten menetelmien avulla voidaan tarkastella tekijöitä, jotka ovat melko yksiselitteisesti luokiteltavissa. Monimutkaisille asioille (esim. onnellisuus, luonteenpiirteet) on vaikea tehdä sellaisia mittareita, joista ei synny erimielisyyttä. Yhdyssanan rakenteen valintaan vaikuttavat rakennepiirteiden keskinäisten suhteiden lisäksi monet pragmaattiset ja semanttiset syyt. En käsittele näitä syitä tässä tutkimuksessa tilastollisesti, mutta varsinkin harvinaisten rakenteiden kvalitatiivisessa analyysissä kiinnitän huomiota myös niihin.

Tutkimuksen kaikki tilastollinen kuvailu ja testaus on tehty R-ohjelmointiympäristön versiolla 3.1.1 (R 2014) RStudio-ohjelmiston (RStudio 2009–2013) kautta. R:n peruspaketin lisäksi olen käyttänyt aineiston käsittelyssä ja testauksessa apuna seuraavia paketteja (versionumero suluissa): *lawstat* (2.4.1), *party* (1.0-15), *plyr* (1.8.1). Tutkimuksen kuviot on tulostettu RStudiosta R:n peruspaketilla, mutta kuvioiden ulkoasu on viimeistelty Adobe Illustrator CS5 -ohjelmalla.

Suurin osa tässä tutkimuksessa käytetyistä tilastollisista tunnusluvuista on yleisesti tunnettuja. Tärkeimmät käytetyt tunnusluvut ovat keskiarvo, mediaani ja keskihajonta. Luvussa 6.2 käytetään kuitenkin myös kahta harvemmin nähtyä tilastollista tunnuslukua: rakenteen sane/lemma-suhdelukua (SLS-luku) sekä rakenteen suhteellista entropiaa (H_{rel}). Koska tunnuslukuja käytetään vain kyseisessä luvussa, ne on esitelty tarkemmin ko. luvun ensimmäisessä alaluvussa (luku 6.2.1). Tutkimuksen muut tilastomenetelmät on esitelty tässä luvussa.

Olen valinnut tilastollisen testauksen merkitsevyytasoksi yleisesti käytössä olevan $\alpha = 0,05$, joka käytännössä tarkoittaa testien 95 %:n luotettavuutta. Useimmissa tutkimuksen testeissä *p*-arvo eli havaittu merkitsevyytaso⁵⁶ on kuitenkin vielä lähempänä nollaa. Raportoinnissa melkein merkitsevät⁵⁷ tulokset ($0,01 < p \leq 0,05$ eli tilanne on 1–5 %:n todennäköisyydellä sattuman aikaansaama) on merkitty

55 Koska yhdyssana-aineiston luotettava automaattinen käsittely ei ole toistaiseksi mahdollista vaan aineisto on annotoitava käsin, otoskokoa ei voida kuitenkaan kasvattaa kovin suureksi. Koska monet marginaaliset tyypit ilmaantuvat otokseen vasta kun sen koko on jo varsin suuri, esimerkiksi A. Hakulisen, Karlssonin ja Vilkun (1996 [1980]: 104–109) esittelemä ajatus juoksevien frekvenssien ”jäätymispisteistä” ei sovellu yhdyssana-aineiston koostamiseen.

56 Tässä tutkimuksessa havaitun merkitsevyytason lyhenteenä on käytetty *p*-kirjainta (*possibility*). Suomenkielisessä tutkimuksessa käytetään toisinaan myös lyhennettä *hmt*.

57 *Melkein merkitsevä* ja *lähes merkitsevä* ovat harhaanjohtavia mutta yleisesti käytettyjä termejä suomenkielisessä tutkimuksessa. Kun tutkimuksen merkitsevyytasoksi on valittu 0,05, myös tulokset, joissa havaittu merkitsevyytaso on $0,01 < p \leq 0,05$, ovat tilastollisesti merkitseviä (eivät ”melkein” merkitseviä). Suomenkielisessä tutkimuksessa voitaisiin vastaavasti käyttää myös termejä *merkitsevä*, *hyvin merkitsevä* ja *erittäin merkitsevä*. Englanniksi vastaavat termit ovat yleensä *significant*, *highly significant* ja *extremely significant*.

yhdeällä tähdeillä (*), merkitsevät tulokset ($0,001 < p \leq 0,01$ eli tilanne on $0,1-1$ %:n todennäköisyydellä sattuman aikaansaama) kahdella tähdeillä (**) ja erittäin merkitsevät tulokset ($p \leq 0,001$ eli tilanne on alle $0,1$ %:n todennäköisyydellä sattuman aikaansaama) kolmella tähdeillä (***)).

Tilastotieteelliset menetelmät ovat tässä tutkimuksessa apuvälineitä, eivät tutkimuksen tai kehittämisen kohteita. Laajan tilastollisiin menetelmiin perehtymisen jälkeen olen valinnut käyttöön pääasiassa perinteisiä ja jo aiemmin tilastollisessa kielitieteessäkin hyväksi havaittuja testejä. Muita menetelmiä olen käyttänyt vain kohdissa, joissa tutkimuksen tavoitteet sitä vaativat. Esittelen seuraavaksi kaikki tutkimuksessa käytetyt tilastolliset testit. Muiden kuin perinteisimpien menetelmien kohdalla perustelen tarkemmin myös niiden valintaa. Ennen jokaisen testin suorittamista olen varmistanut testin taustaoletusten toteutumisen (esim. normaalijakautuneisuus, samahajontaisuus ja riittävä havaintojen määrä kaikissa soluissa). Kaikki tutkimuksessa tehdyt testit ovat kaksisuuntaisia.

Kategoristen muuttujien välillä olevan riippuvuuden tilastollisen merkitsevyyden testauksessa olen käyttänyt ensisijaisesti Pearsonin khiin neliö -riippumattomuustestiä (*Pearson's chi squared test*). Pearsonin khiin neliö -testi on luotettava vain tilanteessa, jossa yhdenkään luokan odotettu frekvenssi ei ole < 1 ja jossa odotettu frekvenssi on < 5 korkeintaan 20 %:ssa ryhmistä. Mikäli nämä oletukset eivät ole kulloinkin tarkasteltavan ominaisuuden kohdalla täyttyneet, olen käyttänyt Pearsonin khiin neliö -testin sijaan Fisherin tarkkaa testiä (*Fisher's exact test*), joka ei ole herkkä pienille solufrekvensseille. (Pearsonin khiin neliö -testistä ja Fisherin tarkasta testistä ks. tarkemmin esim. Pearson 1900, Agresti 2002: 22–24, 91–101). Riippuvuuden vahvuuden ja suunnan tarkastelussa olen käyttänyt lisäksi sovitettuja standardoituja jäännöksiä (*adjusted standardized residuals*), jotka ilmaisevat oletettujen frekvenssien (tilanne, jossa ryhmät eivät eroa toisistaan) ja havaittujen frekvenssien erotuksen. Jäännösten standardointi mahdollistaa erikokoisten ryhmien vertailun. Tässä tutkimuksessa pidän muuttujien välistä yhteyttä tilastollisesti merkitsevänä, kun jäännös on pienempi kuin -2 tai suurempi kuin 2 (ks. Agresti 2002: 80–82).

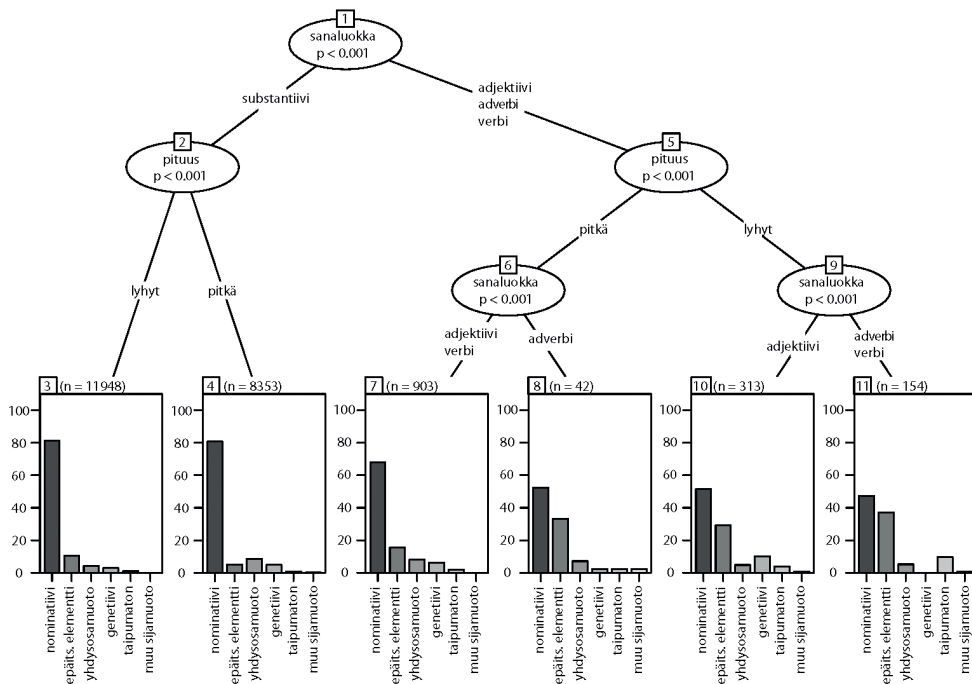
Numeeristen muuttujien normaalijakautuneisuuden testauksessa olen käyttänyt Shapiro–Wilkin normaalijakaumatestiä (*Shapiro–Wilk normality test*, testiä tarkemmin Shapiro–Wilk 1965). Testien perusteella kaikki tutkimuksessa tarkastellut numeeriset muuttujat poikkeavat tilastollisesti merkitsevästi normaalijakaumasta. Numeeristen muuttujien jakaumista piirretyt histogrammit (ks. esim. kuvio 2 sivulla 105) osoittavat, että kaikki jakaumat ovat todellisesta kielestä poimituille aineistoille tyypilliseen tapaan oikealle vinoja ja pitkähäntäisiä (Zipf 1965 ja esim. McEnery–Hardie 2012: 51–52; Kurumada–Meylan–Frank 2013: 440). Varsinkin jakaumien oikeassa laidassa on myös runsaasti poikkeavia havaintoja. Koska kaikki jakaumat ovat samaan tapaan vinoja ja koska käytettävissä on suuri aineisto, aineiston normaalijakautumattomuus ei estä parametristen testien käyttöä. Toinen numeeristen muuttujien tarkastelussa käytettävien testien valinnan kannalta tärkeä taustatieto on tarkasteltavien ryhmien samahajontaisuus. Jakaumien vinouden

vuoksi olen käyttänyt samahajontaisuuden testauksessa mediaaneihin perustuvaa Brownin–Forsythen testiä (*Brown–Forsythe test of variances*, testistä tarkemmin Brown–Forsythe 1974).

Numeerisia muuttujia koskevassa testauksessa olen käyttänyt tilanteen mukaan joko kahden riippumattoman otoksen t-testiä (*independent two-sample t-test*) tai toisistaan riippuvaisten otosten parittaista t-testiä (*paired two-sample t-test*, testeistä tarkemmin Student 1908). Mikäli testattavien muuttujien samahajontaisuusoleus ei ole toteutunut, olen käyttänyt tavanomaisen t-testin sijaan Welchin t-testiä (*Welch's t-test*, testistä tarkemmin Welch 1947). T-testien lisäksi olen käyttänyt tutkimuksessa numeeristen muuttujien erojen tarkasteluun myös monivertailumenetelmiä. Kaikissa tilanteissa, joissa monivertailumenetelmiä on tässä tutkimuksessa tarvittu, vähintään yhden testattavan ryhmän hajonta on poikennut muista tilastollisesti merkitsevästi. Tämän vuoksi monivertailuissa ei ole voitu käyttää tavanomaisia varianssianalyysijä (ANOVA). Se, että ryhmät eivät ole samahajontaisia, tekee periaatteessa myös Kruskal–Wallisin epäparametrisestä varianssianalyysistä epäluotettavan. ANOVAn ja Kruskal–Wallisin testien sijaan olenkin käyttänyt testauksessa R:n *oneway.test*-funktiota (*Test for Equal Means in a One-Way Layout*), joka mahdollistaa Welchin korjauksen käyttämisen myös monivertailuissa (Stats s. v. *oneway.test*; korjauksesta tarkemmin Welch 1951). Welchin korjauksen ansiosta myös keskihajonnaltaan erilaisten ryhmien vertailusta saadaan luotettavaa. *Oneway.test*-funktion heikkous on se, että R:ssä ei ole tarjolla sen tulosten jatkotarkasteluun sopivaa monivertailutestiä. Tämän vuoksi olen tehnyt jatkotarkastelut yksinkertaisesti parittaisten kahden riippumattoman otoksen t-testien avulla.

Tutkimuksen laajimmat monimuuttujamallit olen tehnyt luokittavien puiden avulla (luokittavista puista tarkemmin esim. Breiman ym. 1984; Strobl–Malley–Tutz 2009; kielitieteellisessä tutkimuksessa luokittavia puita ovat käyttäneet esim. Akama 2003; Ernestus–Baayen 2003, 2004; Mendoza–Denton–Hay–Jannedy 2003; Priiki 2014). Luokittavat puut jakavat aineiston osiin, jotka sisältävät keskenään samankaltaisia tapauksia. Jo yksinkertaisella binäärisellä puulla voidaan mallintaa monimutkaisiakin suhteita, sillä puu tekee binäärisiä jakoja rekursiivisesti: se jakaa aineiston ensin kahtia ja sitten jälleen osat kahtia ja niin edelleen.

Nimenomaan luokittavien puiden valintaan tutkimuksessa käytettäväksi monimuuttujamenetelmäksi on kaksi tärkeää syytä: 1) Luokittavat puut ovat muita monimutkaisia tilastomalleja havainnollisempia. Niiden graafisena esitystapana käytetyt puukuvaimet kuvaavat aineiston rakennetta havainnollisesti ja auttavat tutkijaa kiinnittämään huomionsa niihin kohtiin, joita on syytä tarkastella lähemmin muilla menetelmillä (muut tilastolliset menetelmät sekä kvalitatiivinen tarkastelu). 2) Luokittavat puut ovat taustaoletustensa kannalta hyvin suvaitsevaisia. Lisäksi luokittaviin puihin voidaan vaivatta yhdistellä kategorisia ja numeerisia muuttujia niin selitettäväksi kuin selittäviksikin muuttujiksi. Havainnollistan seuraavaksi luokittavaa puuta esittävän kuvan lukemista analogialla luonnollisiin puihin (kuvio 1). Sen jälkeen palaan vielä tarkemmin luokittavien puiden hyviin ja huonoihin puoliin.



KUVIO 1. Esimerkki luokittavasta puusta

Kuviossa 1 on esimerkki luokittavasta puusta (sama puu käydään yksityiskohtaisemmin läpi luvussa 7.2.1 sivulta 196 alkaen). Kuviossa on selitettävänä muuttujana yhdyssanan alkuosan muoto ja selittävinä muuttujina koko yhdyssanan sanaluokka ja pituus. Reaalimaailman puihin nähden luokittava puu on ylösalaisin. Puun lukeminen aloitetaan ylhäältä ensimmäisestä oksanhaarasta eli noodista 1. Puun oksat ja oksanhaarat eli noodit kuljettavat mukanaan tutkimusaineistoa ja sen osia. Kukin oksanhaara jakaa aineiston kahtia yhden selittävän muuttujan perusteella, kunnes päädytään kuviossa alimpana oleviin lehtinoodeihin (pylväsdiagrammeihin). Lehtinoodien diagrammit kuvaavat selitettävän muuttujan (kuviossa 1 alkuosan mahdollisten muotojen) jakaumaa siinä kategoriassa, johon oksia ja niiden haaroja seuraamalla päädytään. Esimerkiksi kuviossa 1 noodin 1 otsikko osoittaa, että noodi jakaa aineiston kahtia alkuosan sanaluokan mukaan. Noodien 1 ja 2 välillä kulkevassa oksassa ovat mukana muista sanaluokista alkuosan muotojakaumaltaan tilastollisesti merkitsevästi ($p < 0,001^{***}$) poikkeavat yhdyssubstantiivit. Noodien 1 ja 5 välillä kulkevassa oksassa puolestaan kulkevat yhdysadjektiivit, -adverbit ja -verbit. Molemmassa oksassa ovat mukana sekä pitkät että lyhyet yhdys sanat, sillä pituuteen liittyvät jaot tehdään vasta myöhemmissä oksanhaaroissa. Esimerkiksi lehtinoodi 10 kuvaa selitettävän muuttujan jakaumaa joukossa, johon kuuluvat yhdys sanat ovat lyhyitä (oksa noodien 5 ja 9 välillä) yhdysadjektiiveja (oksa noodien 9 ja 10 välillä).

Olen valinnut luokittavien puiden kasvattamiseen R:n party-paketin (versio 1.0-15). Party-paketin *ctree*-funktiolla luotujen luokittavien puiden etu moniin muihin samantyyppisiä mallinnuksia tuottaviin paketteihin on se, että *ctree*-puu ei vaadi karsimista tai ristiinvalidointia jälkikäteen (Party s. v. *ctree*; ks. myös Strobl–Hothorn–Zeileis 2009). Ctreessä tutkija voi itse asettaa *p*-arvon, joka jokaisen malliin tulevan binäärisen jaon on alitettava. Tässä tutkimuksessa kaikkien testien $\alpha = 0,05$, joten puissa aktuaalistuvat oksanhaaroiksi vain sellaiset jaot, joissa oksien välinen ero on tilastollisesti merkitsevä vähintään tasolla 0,05 ($p < 0,05$). Puun ensimmäinen oksanhaara (noodi 1) on luokituskaavan tärkeimmäksi valitsema jako eli se, jonka *p*-arvo on pienin. Ctree-funktio tekee malliin tulevat jaot oletusarvoisesti niin, että jokaisessa haarautuvassa noodissa on oltava vähintään 20 havaintoa ja jokaisessa haarassa vähintään 7 havaintoa. Oletusarvot sopivat hyvin tämän tutkimuksen aineiston mallinnukseen, sillä aineistossa on useita pieniä ryhmiä, jotka poikkeavat selvästi muista.

Tässä tutkimuksessa olen käyttänyt luokittavia puita ensisijaisesti aineiston kuvailuun ja muuttujien välillä olevien yhteyksien etsimiseen, mutta olen kuitenkin ilmoittanut kunkin puun yhteydessä myös sen ennustavuusprosentin. Varsinaisten ennustemallien tekemiseen sopivat yksittäisiä luokittavia puita paremmin aineistosta poimituista useista satunnaisotoksista kasvatetut satunnaismetsät (ks. tarkemmin esim. Lohmann 2013). Yksittäisen puun perusteella tehtävät koko populaatioita koskevat ennustemallit ovat satunnaismetsiä epätarkempia, sillä hierarkkisia jakoja rekursiivisesti tekevässä mallinnuksessa pienetkin erot otoksissa voivat johtaa hyvin erilaisiin lopputuloksiin. (Strobl–Malley–Tutz 2009; Lohmann 2013: 10.) Yksi koko aineistoa kuvaava puu näyttää kuitenkin havainnollisemmin aineiston päälinjat ja sopii siksi satunnaismetsiä paremmin tämän tutkimuksen kuvaileviin tarkoituksiin.

Useiden kategoristen muuttujien välillä olevien yhteyksien tarkasteluun on perinteisesti käytetty logistista regressiota. Logistinen regressio on kuitenkin herkkä muuttujien välillä oleville voimakkailla keskinäisille riippuvuuksille (esim. Eddington 2010; Lohmann 2013: 1). Koska tutkimusaineiston muuttujien välillä on useita voimakkaita keskinäisiä riippuvuuksia, taustaoletuksiltaan suvaitsevaisen menetelmän valinta on tässä tutkimuksessa välttämätöntä. Lisäksi – kuten edellä on jo käynyt ilmi – tutkimuksessa tarkasteltavat numeeriset muuttujat eivät täytä perinteisimpien tilastomenetelmien vaatimuksia kuten normaalijakautuneisuutta ja samahajontaisuutta. Luokittelevat puut tarjoavat turvallisen keinon analysoida vinojakaumaisia numeerisia muuttujia, joissa ryhmien varianssit poikkeavat toisistaan.

Myös tutkimusaineiston kategoriset muuttujat ovat hyvin epätasaisesti jakautuneita. Tällöin suurimmat ryhmät pääsevät helposti dominoimaan aineistosta muodostettuja tilastollisia malleja ja pienten ryhmien ominaisuudet katoavat näkyvistä. Parittaisessa testauksessa ja pienemmissä malleissa ongelma voidaan kiertää käyttämällä suuremmista ryhmistä pienempiä satunnaisotoksia (kuten tässäkin tutkimuksessa on tehty). Kun mallissa on mukana useita (vieläpä hyvin epätasaisesti

jakautuneita) muuttujia, satunnaisotosten käyttö saattaa kadottaa aineistosta jonkin ryhmän kokonaan tai jättää jäljelle niin pieniä ryhmiä, että tuloksia ei voida enää pitää luotettavina (Mendoza-Denton-Hay-Jannedy [2003: 128] tosin huomauttavat, että puuttuvat tiedot tai pienet tai tyhjät solut eivät sinänsä ole ongelma luokittelevalle puuanalyysille). Havainnollisuutensa ansiosta luokittava puu selviää monia muita mallitusmenetelmiä paremmin myös epätasaisesti jakautuneiden kategoristen muuttujien käsittelystä: puukuvaimen ansiosta suurten ryhmien ”kulkua” mallissa voidaan seurata koko mallin tulkinnan ajan. Samalla nähdään, minkä muiden ryhmien tuloksiin suuret ryhmät mahdollisesti vaikuttavat ja mitkä ryhmät kulkevat jo omissa oksissaan riippumattomina dominoivista suurista ryhmistä. Mallin havainnollisuuden ansiosta tutkija voi tapauskohtaisesti harkita, onko suuren ryhmän ja sen mallissa mukanaan kuljettamien pienten ryhmien eroa testattava vielä parittaisesti.

Eri muuttujista luokittavat puut kuvaavat havainnollisimmin puumallissa ensimmäisenä olevan selittävän vaikutusta selitettävän muuttujan jakaumaan. Ensimmäisen oksanhaaran jälkeen analyysiin mukaan tulevista muuttujista sen sijaan voi olla vaikea saada yleiskuvaa. Tässä tutkimuksessa luokittavien puiden havainnollisuutta muiden selittävien muuttujien osalta on parannettu tekemällä useita puumalleja, joissa ovat mukana eri muuttujat ja aineiston eri osat (ks. myös luku 7.2).

Mitä enemmän tutkimuksessa tehdään monivertailuja ja parittaisia vertailuja, sitä suuremmaksi kasvaa riski, että jokin löydetystä tilastollisesti merkitsevistä eroista on seurausta puhtaasta sattumasta. Tämän riskin vuoksi kaikki tässä tutkimuksessa esitetyt monivertailujen ja parittaisten vertailujen p-arvot on korjattu Bonferronin–Holmin korjauksella (*Bonferroni–Holm correction for multiple comparisons*). Bonferronin–Holmin korjaus on konservatiivisena tunnettuun Bonferronin korjaukseen perustuva p-arvojen korjausmenetelmä, joka perinteiseen Bonferronin korjaukseen verrattuna vähentää II tyypin⁵⁸ virheen mahdollisuutta (Holm 1979). Tutkimuksessa ainoastaan luokittavien puiden p-arvot on korjattu perinteisellä Bonferronin korjauksella, sillä puiden kasvattamiseen käyttämäni *ctree*-funktio ei tarjoa mahdollisuutta Bonferronin–Holmin korjauksen käyttöön (Party s. v. *ctree*). Molemmissa korjauksissa p-arvot on korjattu samaa kysymystä tarkastelevien testien määrällä eikä esimerkiksi kaikkien tutkimuksessa tehtyjen testien määrällä.

58 Konservatiivisen Bonferronin korjauksen avulla voidaan yleensä välttää hyvin tyypin I virheet, joissa nollahypoteesi hylätään vaikka se todellisuudessa pitääkin paikkansa (ts. esim. muuttujien välille oletetaan yhteys, vaikka niiden välillä ei todellisuudessa ole yhteyttä). Konservatiivisen korjauksen käyttö lisää kuitenkin riskiä tyypin II virheeseen, jossa nollahypoteesia ei hylätä, vaikka se ei pidä paikkaansa (ts. esim. muuttujien välille ei oleteta yhteyttä, vaikka yhteys todellisuudessa on olemassa; virhetyypeistä esim. Luojola 2006: 90–92). Eräiden tutkijoiden mukaan p-arvojen korjauksia ei tulisi tästä syystä käyttää lainkaan (esim. Rothman 1990; Gelman–Hill–Yajima 2008). (Katsaus aiheeseen esim. Feise 2012.)

5 Yhdyssanojen rakennepiirteiden jakauma

Tässä luvussa käsittelen uusien yhdyssanojen rakennetta ominaisuus kerrallaan. Luvun pääasiallinen menetelmä on kvantitatiivinen kuvailu, jota tukee kvalitatiivinen tarkastelu. Vertailussa aiempiin tutkimuksiin olen käyttänyt myös tilastollisia testejä, joiden avulla olen arvioinut havaittujen erojen tilastollista merkitsevyyttä. Tästä luvusta eteenpäin kaikki tekstissä mainitut yhdyssanaesimerkit ovat omasta tutkimusaineistostani (TT), ellei toisin ole mainittu. Selvyyden vuoksi joissain kohdissa on käytetty edelleen myös TT-lyhennettä.

Ensimmäisessä alaluvussa (5.1) käsittelen yhdyssanan pituuteen liittyviä piirteitä: leksikaalisten osien määrää sekä koko muodosteen ja sen osien merkkimäärää. Toisessa alaluvussa (5.2) tarkastelen yhdyssanoihin ja sanaluokkiin liittyviä seikkoja: koko muodosteen ja sen osien sanaluokkaa. Kolmannessa alaluvussa (5.3) kuvaan yhdyssanan alkuosan muotoa: alkuosan morfologista muotoa sekä sen kongruenssia yhdyssanan jälkiosan kanssa.

5.1 Pituus

Tässä luvussa selvitän, kuinka pitkiä suomen kirjoitettuun yleiskieleen syntyvät uudet yhdyssanat ovat sekä leksikaalisten osien (5.1.1) että merkkien (5.1.2) määrällä mitattuna. Produktiivisuuden kannalta tämä on tärkeää, sillä näin päästään tarkastelemaan pituuden produktiivisuudelle asettamia rajoja.

5.1.1 Leksikaalisten osien määrä

Yhdyssanan osina voivat toimia sekä itsenäiset sanat että erilaiset epäitsenäiset leksikaaliset elementit (ks. tarkemmin luku 2.2.3). Käytän tässä luvussa ne kaikki kattavaa nimitystä *leksikaalinen osa*⁵⁹ Leksikaalisen osan käsite on huomattava erottaa yhdysosan käsitteestä. *Yhdysosa*-termillä viitataan tässä tutkimuksessa ainoastaan yhdyssanan semanttisiin rakennesein. Aina näitä käsitteitä ei ole tutkimuksessa tarkkaan erotettu (esim. ISK § 403), vaikka se selvyyden vuoksi olisi tarpeen.

59 Esimerkiksi Koivisto (2013: 256–258) käyttää leksikaalisista osista nimitystä (*yhdyssanan*) *osana*.

Esimerkiksi yhdyssanan *lähiulkomaa* yhdysosat ovat määriteosa *lähi-* ja edusosa *-ulkomaa*. Saman yhdyssanan leksikaaliset osat puolestaan ovat *lähi*, *ulko* ja *maa*. Näistä *maa* on itsenäinen sana ja *lähi* ja *ulko* epäitsenäisiä leksikaalisia elementtejä. Lyhyiden vuoksi viittaa tässä luvussa tietyn määrän leksikaalisia osia sisältäviin yhdyssanoihin myös senkaltaisilla nimityksillä kuin kaksiosainen, kolmiosainen ja niin edelleen.

Tässä luvussa on käytetty osa-aineistoa 1, joka sisältää yhden kirjoitusasun kaikista aineiston yhdyssanoista ($n = 28\ 013$; osa-aineistot on esitelty tarkemmin luvussa 4.3). Tiedot tutkimusaineiston yhdyssanojen leksikaalisten osien määrästä on koottu taulukkoon 1.

TAULUKKO 1. Tutkimusaineiston yhdyssanojen jakauma leksikaalisten osien määrän mukaan ($n = 28\ 013$)

Osien määrä	Esimerkki	Osuus	Määrä
2	<i>haamuasiakas, rehellisyysvaje</i>	83,6 %	23 410
3	<i>hirvikärpäshousut, koodinetsintäohjelma</i>	15,5 %	4 347
4	<i>ammattikorkeakoulutohtori, sähkökirjaliikevaihto</i>	0,9 %	248
5	<i>kehityksmaasairaanhoidajakoulutus, lähialueministerivaliokunta</i>	0,02 %	7
6	<i>antakaa meidän olla rauhassa -yhteiskunta</i>	0,004 %	1
Yhteensä		~ 100 %	28 013

Kuten taulukosta 1 nähdään, ylivoimaisesti suurin osa uusista yhdyssanoista koostuu kahdesta leksikaalisesta osasta. Kaksiosaisien yhdyssanojen osuus on tutkimusaineistossa 83,6 %. Kolmiosaisia yhdyssanoja on aineistossa 15,5 % ja neliosaisia 0,9 %. Tätä useampiosaisia sanoja on aineistossa vain muutamia (viisiosaisia 7 kpl ja kuusiosaisia 1 kpl). Seitsenosaisia tai sitä pidempiä tapauksia aineistossa ei ole lainkaan. Aineiston yhdyssanojen keskipituus on 2,17 leksikaalista osaa ja mediaani 2 osaa.

Taulukossa 1 esitelty jakauma osoittaa, että suomen kielen uusissa yhdyssanoissa on vain hyvin harvoin enemmän kuin 4 leksikaalista osaa. Aiemmissä tutkimuksissa on esitetty, että suomen kielen yhdyssanoissa on harvoin enemmän kuin 5 leksikaalista osaa (Multanen 1991: 61; ISK § 405; Lantee 2010: 33; ISK:n tiedot perustuvat mahdollisesti Multasen tutkimukseen; ISK:n lähdeviitteiden kritiikkiä esim. Häkkinen 2005: 558–559). Oma tutkimukseni kuitenkin osoittaa, että myös 5-osaiset yhdyssanat ovat hyvin harvinaisia eikä uusia 4-osaisiakaan yhdyssanoja synny paljon. Tutkimusaineiston ainoa 6-osainen yhdyssana on ketjumainen *antakaa meidän olla rauhassa -yhteiskunta*, jota on kontekstissaan käytetty selittämässä lyhennettä, jolla samaan asiaan viitataan jatkossa (esim. 9). Pitkä ketjuyhdyssana olisikin hankala toistuvassa käytössä.

- (9) Ilman yhteisöllisyyteen ja kansalaisuuteen kasvattavia vastatoimia uhkana on amor- eli *antakaa meidän olla rauhassa -yhteiskunta*. – Amor-yhteiskunnan lempeä kutsu näkyy jo nyt muun muassa äänestysinnon vähenemisenä. [TT <- HS 24.7.2009, Hannu Kareinen, s. A 2]

Aineistoni 5-osaisista yhdyssanoista kaikki 7 ovat substantiiveja. Yksikään niistä ei sisällä ketjumaisia osia. 4-osaisista yhdyssanoista (248 kpl) ketjumainen osuus sisältyy kahdeksaan yhdyssanaan (mm. *pyydä ja päästä -kalastaja, tee-se-it-se-genetiikka*). Sekä 4- että 5-osaisissa yhdyssanoissa yleisiä rakenneosia ovat sellaiset lyhyet ja hyvin leksikaalistuneet yhdyssanat, joita tavalliset kielenkäyttäjät eivät välttämättä ensisilmäyksellä miellä yhdyssanoiksi (esim. *elokuvateatterielokuva, ikkunalaunapuutarha, koulukirjaesimerkki*). Vaikka tällaiset yhdyssosat selvästi koostuvat kahdesta leksikaalisesta elementistä, ne todennäköisesti hahmotetaan ja tulkitaan yleensä kokonaisuutena (yhdyssanojen tulkinnasta ks. tarkemmin teori-aluku 2.3.3). Tällaisia osia sisältävät pitkät yhdyssanat ovat todennäköisesti helpommin tulkittavia ja siten käytössä keveämpiä kuin sellaiset moniosaiset yhdyssanat, joihin ei sisälly vanhastaan leksikaalistuneita yhdyssanoja (esim. 5-osainen *kesto-magneettitahtikonekäyttö* ja 4-osaiset *eläingeenivaraohjelma* ja *mikrovirtakohotushoito*). Kaiken kaikkiaan yhdyssanan leksikaaliset osat näyttävät olevan sitä lyhyempiä, mitä enemmän osia on. Tarkastelen tätä tendenssiä tarkemmin yhdyssanan rakennepiirteiden keskinäisiä suhteita käsittelevässä luvussa 7.1.1.

Suurin osa uusista pitkistä yhdyssanoista on substantiiveja. Nimenomaan yhdyssubstantiivien leksikaalisten osien määrää on aiemmin tutkinut Eeva Multanen (1991) tilapäisten yhdyssubstantiivien semantiikkaa käsittelevässä pro gradu -tutkielmassaan. Multasen aineisto käsittää 988 vuonna 1989 lehtikielestä poimitua tilapäiseksi tulkittua yhdyssubstantiivivia (mts. 5–8, 16). Tiedot oman aineistoni ja Multasen aineiston yhdyssubstantiivien pituudesta on esitetty taulukossa 2.

TAULUKKO 2. Tutkimusaineiston ja Multasen (1991) aineiston yhdyssubstantiivien jakauma leksikaalisten osien määrän mukaan

Osien määrä	Uudet yhdyssubstantiivit vuosilta 2000–2009 ($n = 26\ 004$)	Tilapäiset yhdyssubstantiivit vuodelta 1989 (Multanen 1991, $n = 988$)
2	82,8 %	78,3 %
3	16,2 %	19,4 %
4	0,9 %	1,7 %
5	0,03 %	0,6 %
6	0,004 %	0,0 %

Kuten taulukosta 2 nähdään, 2-osaisten yhdyssubstantiivien osuus on omassa tutkimusaineistossani suurempi kuin Multanen (1991: 61) aineistossa. Sekä 3-, 4- että 5-osaisten yhdyssanojen osuus puolestaan on selvästi pienempi kuin Multasella. Oman aineistoni yhdyssubstantiivien keskipituus on 2,18 leksikaalista osaa. Multasen aineiston yhdyssubstantiivien keskipituus puolestaan on 2,24 leksikaalista osaa (laskettu Multasen raporttoimista tiedoista tätä tutkimusta varten). Vuoden 1989 tilapäiset substantiivit ovat siis leksikaalisten osien määrällä mitattuna keskimäärin hieman pidempiä kuin vuosien 2000–2009 uudet yhdyssanat ylipäänsä (yhdyssanan tilapäisyydestä ja vakiintumismahdollisuuksista ks. esim. Downing 1977: 818–824). Ero on tilastollisesti merkitsevä (riippumattomien

otosten t-testin $t = -3,24$, $n = 1976$, $df = 1903$, $p = 0,003^{**}$). Ero voi olla ainakin osittain seurausta aineistojen erilaisesta luonteesta (tilapäiset yhdyssanat / uudisyhdyssanat), eikä tulos siksi varmasti osoita, että suomen kirjoitetun yleiskielen uudet yhdyssanat olisivat lyhentyneet viimeisten kymmenien vuosien kuluessa. Tulos on kuitenkin sikäli huomionarvoinen, että vielä 1900-luvun loppupuolella suomen kielen sanojen arveltiin ennemmin pidentyneen kuin lyhentyneen (esim. Pulkkinen 1972: 74–78). Myös kompleksisten muodosteiden – kuten yhdyssanojen – ja pitkien sanojen käytön esitettiin lisääntyneen monissa tekstilajeissa (Rajakallio 1995: 38; Niemikorpi 1996: 9).

Pauli Brattico, Alina Leinonen ja Christina M. Krause (2007) ovat osoittaneet suomen kielen johtamista koskevassa kokeellisessa tutkimuksessaan, että sanamuodostuksen sääntöjen lisäksi myös morfologinen kompleksisuus vaikuttaa siihen, kuinka hyväksyttävänä ihmiset pitävät johdoksia. Bratticon, Leinosen ja Krausen (mts. 126–131) mukaan kielenkäyttäjät arvioivat keksityt mutta sinänsä kieliopilliset johdokset sitä todennäköisemmin kielenvastaisiksi ja merkityksettömiksi, mitä kompleksisempia johdokset ovat eli mitä useampia johtimia ne sisältävät. Esimerkiksi sanaa *piste|ke* (2 morfia, alkuperäisessä artikkelissa engl. *morpheme*) pidetään todennäköisemmin kielenmukaisena ja merkityksellisenä kuin sanaa *apina|lis|ta|tta|ma|ssa*⁶⁰ (6 morfia) (esimerkit mts. 132–135). Yli kuusi morfia sisältävät johdokset eivät ole suomen kielessä mahdollisia, sillä kielenkäyttäjät pitävät niitä niiden kompleksisuuden vuoksi kielenvastaisina vaikka ne muuten noudattaisivat johto-opin sääntöjä. Kuuden morfin raja ei kuitenkaan ole yhtä ehdoton kuin johto-opin säännöt, sillä epäkieliopilliset johdokset arvioidaan kielenvastaisiksi ja merkityksettömiksi niiden kompleksisuudesta riippumatta.

Yhdyssanojen muodostuksen ajatellaan olevan luonteeltaan rekursiivista, eikä yhdyssanassa esiintyvien leksikaalisten osien määrälle pitäisi olla mitään periaatteellista ylärajaa. Käytännössä produktiivisuuden raja näyttää yhdyssanojen muodostuksessa kuitenkin asettuvan noin viiden leksikaalisen osan⁶¹ kohdalle (ks. edellä). Vaikka harvoissa uusissa yhdyssanoissa on enemmän kuin 4 leksikaalista osaa, hyvinkin pitkiä yhdyssanoja voidaan pitää sekä kielenmukaisina että ymmärrettävinä. Esimerkkinä mainittakoon suomen kielen pisimmäksi todellisessa käytössä olevaksi sanaksi väitetty *lentokonesuihkurbiinimoottoriapumekaanikkoliupseerioppilas*, joka sisältää 10 leksikaalista osaa. Yhdyssanojenkin kohdalla ymmärrettävyyden ja kieliopillisuuden raja tulee varmasti jossain kohtaa vastaan,

60 Brattico, Leinonen ja Krause (2007) eivät tuo esiin morfijaossa käyttämiään periaatteita eikä morfien rajoja ole merkitty näkyviin tutkimusaineiston sanoihin. Ainoastaan morfien määrä kussakin sanassa on ilmoitettu. Esimerkiksi Bratticon ym. kuusimorfiseksi merkitsemässä sanassa *apinallistatamassa* yhdeksi morfiksi on laskettu myös yhdysperäinen *-linen-*johdin (yhdysperäisyydestä ks. L. Hakulinen 1979: 164–167) sekä inessiivin sijapäätte *-ssa*, joka kuuluu enemmän taivutus- kuin johtomorfologiaan. Tässä mainittuihin esimerkkeihin olen merkinnyt morfien rajat itse.

61 Esimerkiksi viidestä leksikaalisesta osasta koostuva yhdyssana voi tietysti sisältää paljon enemmän morfeja, sillä myös johdokset voivat toimia yhdyssanan osana. Johdinten merkitys on yleensä epätarkempi kuin itsenäisten lekseemien ja leksikaalisilla elementeillä, ja johtimilla on siten erilainen vaikutus sanan merkityksen rakentumiseen.

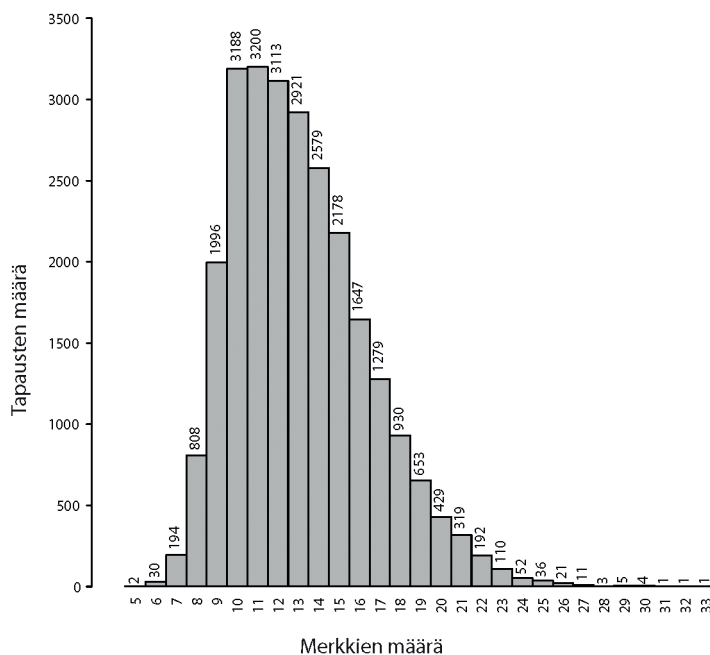
mutta todellisessa kielenkäytössä paljon ennen sitä alkaa rakenteen produktiivisuutta rajoittaa ilmauksen käytettävyys.

Pitkät yhdyssanat ovat käytössä monella tapaa hankalia. Koska yhdistämällä sanoja yhdyssanoiksi voidaan tiivistää paljon tietoa pieneen tilaan ja koska sanojen yhdistäminen yhdyssanaksi usein hävittää tiedot osien syntaktisista suhteista, varsinkin pitkän yhdyssanan tulkitseminen voi olla vaikeaa ja raskasta. Pitkän yhdyssanan tulkitseminen vaatiikin usein enemmän kognitiivisia ponnisteluja kuin syntaktisen rakenteen tulkitseminen (esim. *jäsenmaksuvapautusperuste ~ peruste jäsenmaksusta vapauttamiseen*, Iisa–Piehl 1992: 79–81; ks. myös Inhoff–Radach–Heller 2000; Hyönä–Vainio 2009: 725–726). Varsinkin useasta perättäisestä nominatiivimuotoisesta sanasta koostuvan yhdyssanan merkitys voi olla vaikea hahmottaa. Uuden yhdyssanan tulkitseminen vaatii usein tuekseen myös tiedon kontekstista (Štekauer 2009; ks. myös luku 2.3.1). Tulkitsemisvaikeuksien lisäksi pitkät yhdyssanat voivat aiheuttaa ongelmia myös tekstin asettelussa esimerkiksi lehden taittoprosessissa. Pitkiä sanoja on myös pidetty suorastaan epäesteettisinä. Näiden syiden vuoksi kielenhuolto onkin jo pitkään ohjeistanut välttämään liian pitkiä yhdyssanoja ja tulkinnanvaraisissa tapauksissa kirjoittamaan ilmauksen mieluummin sanaliitoksi kuin yhdyssanaksi (Kanervamäki 1996: tiivistelmä ja 18–20, 96–97; erikoiskielten osalta ks. Haarala 1989: 271). Myös termityössä suositaan mieluummin lyhyitä kuin pitkiä sanoja, ja esimerkiksi Sanastokeskus TSK:n tietotekniikan termi-talkoiden ohjeissa neuvotaan välttämään neliosaisia ja sitä pidempiä yhdyssanoja ja sanaliittoja (TSK: Millä perusteella termisuosituksia; ks. myös Birdlife: Periaatteet).

5.1.2 Merkkimäärä

Tässä luvussa käsitelen uusien yhdyssanojen pituutta sekä kokonaisten yhdyssanojen että niiden osien merkkimäärän kannalta. Eräs kokonaisten yhdyssanojen osalta mielenkiintoinen tutkimuskysymys on, onko ihmisen tarkan näön alueelle mahtuvien merkkien määrällä yhteyttä uusien yhdyssanojen pituuteen. Koska ihmisen tarkan näön alueelle mahtuu vain noin 12 merkkiä kerrallaan, ihminen tarvitsee sitä pidempien sanojen lukemiseen yleensä enemmän kuin yhden silmän fiksaation. Fiksaatioiden lisääntyminen puolestaan lisää todennäköisyyttä sille, että sana prosessoidaan osissa. (Ks. tarkemmin luku 2.3.3 sekä Bertram–Hyönä 2003). Yhdyssanan osien kannalta oleellinen kysymys on yhdyssanan alku- ja jälkiosan keskimääräinen pituus. Koska sanan pituus on suomen kielessä yleensä yhteydessä sen morfologiseen kompleksisuuteen, viittäisi jommankumman keskimääräinen pidemmyys myös suurempaan morfologiseen kompleksisuuteen.

Luvussa on käytetty osa-aineistoa 2, joka sisältää aineiston kaikki sellaiset yhdyssanat, joissa ei ole välimerkkejä tai lyhenteitä ($n = 25\ 903$; osa-aineistot on esitelty tarkemmin luvussa 4.3). Aineiston yhdyssanojen merkkimääräjakaumaa on havainnollistettu kuviossa 2.



KUVIO 2. Tutkimusaineiston yhdyssanojen merkkimääräjakauma ($n = 25\,903$)

Kuten kuviosta 2 nähdään, aineistossa yleisimpiä ovat 10–12 merkkiä sisältävät muodosteet. 5 ja 10 merkin välillä tapausten määrä kasvaa suurin harppauksin. 12 merkin jälkeen tapausten määrä kääntyy laskuun ja laskee melko tasaisesti, kunnes tasoittuu jakauman pitkäksi hännäksi. Jakauma on positiivisesti eli oikealle vino, eikä se noudata normaalijakaumaa (Shapiron–Wilkin normaalijakaumatestin $p < 0,001^{***}$). Jakaumassa on 437 poikkeavaa havaintoa (kaikki ≥ 22 merkkiä pitkät tapaukset). Jakaumassa on kuitenkin havaintoja kaikissa luokissa välillä 22–33 merkkiä. Jakaumassa näkyvä pitkä oikealle osoittava häntä ja poikkeavat havainnot ovat tyypillisiä luonnollisesta kielestä kootulle aineistolle (Zipf 1965 ja esim. McEnergy–Hardie 2012: 51–52; Kurumada–Meylan–Frank 2013: 440), jossa frekventtejä tapauksia seuraa joukko vähäfrekventtisempiä tapauksia, jollaisten todennäköisyys on populaatiossa pieni. Aineiston jakauma on samaan tapaan oikealle vino myös tarkasteltaessa erikseen 2, 3 ja 4 leksikaalista osaa sisältäviä yhdyssanoja.

Jakauman vinouden ja siinä olevien lukuisien poikkeavien havaintojen vuoksi jakauman tarkastelussa on periaatteessa parempi käyttää robusteja tunnuslukuja, jotka eivät ole herkkiä vinoudelle ja poikkeaville arvoille. Tilastollisista tunnusluvuista aritmeettista keskiarvoa paremmin jakaumaa kuvaa mediaani. Jakauman vinoudesta huolimatta sen mediaani ja keskiarvo ovat kuitenkin lähellä toisiaan; mediaani on 13 merkkiä ja keskiarvo 13,1 merkkiä. Tärkeimmät jakaumaa kuvaavat tilastolliset tunnusluvut on koottu taulukoon 3.

TAULUKKO 3. Tutkimusaineiston yhdyssanojen merkkimääräjakauman tilastolliset tunnusluvut

Tunnusluku	Arvo
havaintojen määrä	25 903
mediaani	13
moodi	11
keskiarvo	13,1
minimi	5
maksimi	33
alakovartiili	11
yläkovartiili	15
keskihajonta	3,33

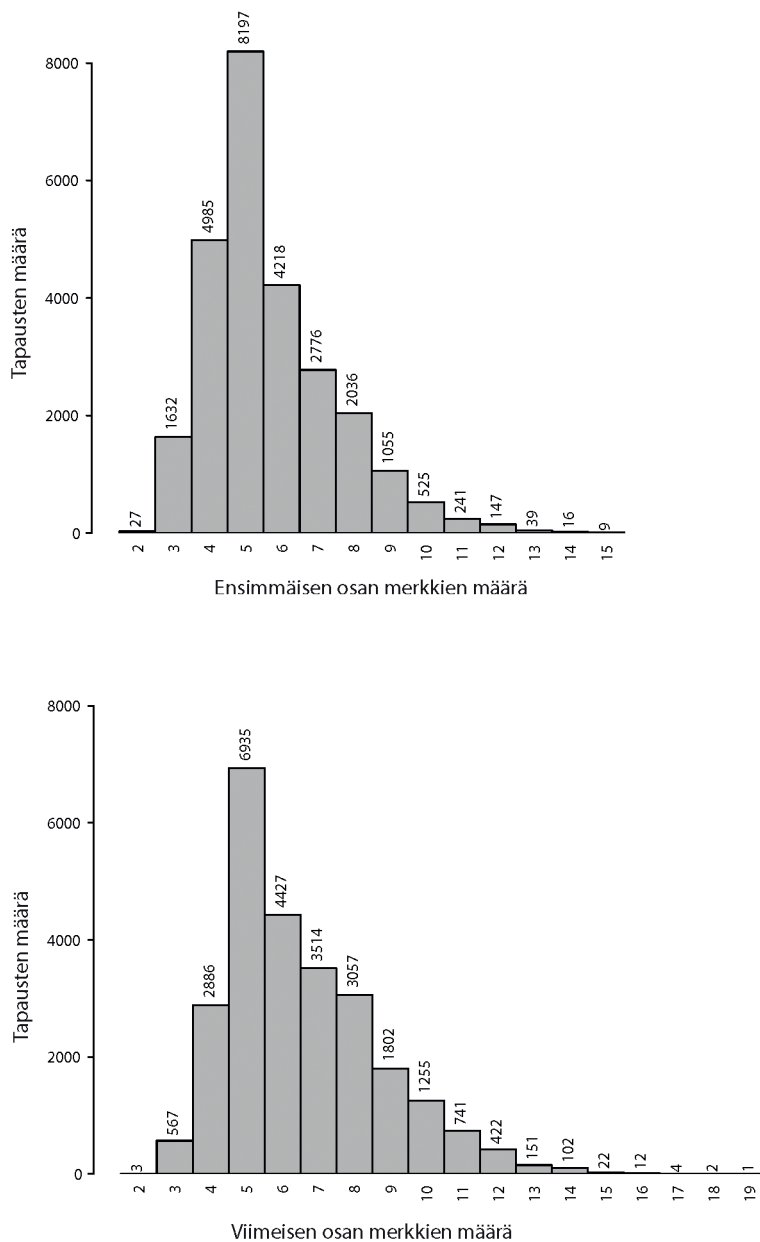
Kuten taulukosta 3 nähdään, aineiston yleisin merkkimäärä eli merkkimäärän moodi on 11 merkkiä (12,4 %, esim. *luonnonvesi, takuupalkka*). 10- ja 12-merkkisiä yhdyssanoja on lähes yhtä paljon kuin 11-merkkisiä (10-merkkisiä 12,3 % ja 12-merkkisiä 12,0 %). Jakauman vinouden vuoksi mediaani ja keskiarvo nousevat kuitenkin 13 merkkiin. Keskihajonta on 3,3 merkkiä. Aineistosta 50 % sisältää 11–15 merkkiä. Aineiston lyhimmat sanat sisältävät 5 merkkiä (*ylivyö* [termistä *ylivyön korko*] ja *häyö* 'yö, jona kirkossa pidetään vihkitilaisuuksia') ja pisin 33 merkkiä (*maailmassaselviytymiskirjallisuus*).

Sitä, että uusien yhdyssanojen produktiivisuuspiikki osuu juuri 10–12 merkkiä sisältävien muodosteiden kohdalle, selittää ainakin kolme seikkaa: 1) Uusia vähemmän kuin 10 merkkiä sisältäviä yhdyssanoja syntyy vain vähän, koska niiden osina käytettävät suomen kielen perussanat ja johdokset ovat jo itsessään melko pitkiä⁶². 2) Pitkät yhdyssanat ovat epäkäytännöllisiä sekä hahmottamisen, tekstin asettelun että muiden seikkojen suhteen. Periaatteessa yhdyssanojen muodostus on kuitenkin rekursiivista, ja pitkiäkin sanoja voidaan tarvittaessa luoda. Käytännössä sananmuodostuksessa pyritään silti lyhyteen. 3) Pitkien sanojen hankala hahmottuminen liittyy ihmisen tarkan näön alueeseen, jolle mahtuu noin 12 merkkiä kerrallaan. Ei välttämättä ole sattumaa, että uusien yhdyssanojen määrä alkaa laskea juuri tarkan näön alueen rajan tullessa vastaan.

Kokonaisten yhdyssanojen pituus on suoraa seurausta niiden osien pituudesta. Yksittäisten osien merkkimäärän tarkempaan tarkasteluun sopivimpia osia ovat yhdyssanojen ensimmäiset ja viimeiset osat (siis 2-osaisissa toinen osa, 3-osaisissa kolmas osa ja 4-osaisissa neljäs osa), sillä ne ovat kaikissa yhdyssanoissa samantyyppisessä asemassa sanan leksikaalisten osien määrästä riippumatta (ks. myös luku

62 Yhdyssanojen, perussanojen ja johdosten keskimääräistä pituutta ei ole tutkittu erikseen, mutta kaikki sanaryhmät käsittävä suomen kielen keskimääräinen sanapituus esimerkiksi lehtikielessä on lähteestä riippuen noin 7,2–9,6 merkkiä (Niemi 1990: 9, 1991: 86; Heikkinen – O. Lehtinen – Lounela 2002). Lemmatun aineiston keskimääräinen sanapituus on jonkin verran pienempi, sillä juoksevässä tekstissä myös muut suffixit kuin johtimet pidentävät saneita.

7.1.1). Aineiston yhdyssanojen ensimmäisten ja viimeisten osien merkkimääräjakamia on havainnollistettu kuviossa 3.



KUVIO 3. Tutkimusaineiston yhdyssanojen ensimmäisen ja viimeisen osan merkkimääräjakauma ($n = 25\,903$)

Kuten kuvioista 3 nähdään, ensimmäisten ja viimeisten osien merkkimääräjakumat ovat samaan tapaan oikealle vinoja kuin koko muodosteen merkkimääräjakumakin (kuvio 2). Viimeisten osien jakaumassa on jonkin verran enemmän hajontaa kuin ensimmäisten osien jakaumassa. Tämä näkyy siitä, että viimeisissä osissa on enemmän pitkiä tapauksia ja jakauman oikealle osoittava häntä jatkuu pidemmälle. Ensimmäisten osien jakaumassa puolestaan tapauksista jopa yli 30 %:n merkkimäärä on sama kuin jakauman moodi (moodi = 5; esim. *kaljaranneke, nettivaate*).

Uusien yhdyssanojen ensimmäisen leksikaalisen osan mediaanipituus on 5 merkkiä (keskiarvo 5,7 merkkiä). Myös yleisin merkkimäärä eli moodi on 5. Keskihajonta on 1,8 merkkiä. Aineiston alkuosista 50 % sisältää 4–7 merkkiä. Lyhimmät alkuosat sisältävät 2 merkkiä (esim. *me-tunne, yöpyöräily*) ja pisimmät 15 merkkiä (esim. *markkinoistamismalli, puuttumattomuussopimus*). Poikkeavia havaintoja on 211 kpl (kaikki ≥ 12 merkkiä sisältävät tapaukset; jakauma on kuitenkin jatkuva välillä 11–15 merkkiä).

Uusien yhdyssanojen viimeisen leksikaalisen osan mediaanipituus on 6 merkkiä (keskiarvo 6,5 merkkiä). Samoin kuin alkuosan, myös jälkiosan yleisin merkkimäärä on 5 (esim. *mummopommi, reittitaksi*). Keskihajonta on 2,1 merkkiä. Aineiston alkuosista 50 % sisältää 5–8 merkkiä. Lyhimmät jälkiosat sisältävät 2 merkkiä (3 yö-sanana esiintymää: *hääyö, sysiyö ja yliyö*) ja pisin 19 merkkiä (*uusbyrokraatisoituminen*). Poikkeavia havaintoja on 294 kpl (kaikki ≥ 13 merkkiä sisältävät tapaukset; jakauma on kuitenkin jatkuva välillä 13–19 merkkiä).

Uusien yhdyssanojen alkuosat ovat tilastollisesti merkitsevästi lyhyempiä kuin jälkiosat (parittaisten otosten t-testin $t = -9,6$, $n = 2\ 000$, $df = 999$, $p < 0,001^{***}$). Alkuosien keskimääräinen lyhemmyys on osittain seurausta lyhyiden prefiksinkaltaisten alkuosien sekä erilaisten lyhentyneiden yhdysosamuotoisten alkuosien käytöstä (yhdysosamuotona voi toimia esim. pelkkä sanavartalo, ks. tarkemmin luku 2.2.3). Ero alku- ja jälkiosien keskimääräisessä pituudessa on kuitenkin tilastollisesti merkitsevä myös verrattaessa pelkästään sellaisia yhdyssanoja, joissa ei ole prefiksinkaltaisia elementtejä tai yhdysosamuotoja (parittaisten otosten t-testin $t = -7,4$, $n = 2\ 000$, $df = 999$, $p < 0,001^{***}$). Koska aineisto on lemmattua, taivutus-päätteet eivät selitä jälkiosien keskimääräistä pidemmyyttä (ks. myös Niemikorpi 1991: 85). Koska sanan pituus on suomen kielessä usein seurausta sen kompleksisuudesta, uusien yhdyssanojen alkuosien keskimääräinen lyhemmyys viittaa siihen, että yhdyssanan alkuosan morfologista kompleksisuutta pyritään suomen sanamuodostuksessa jossain määrin välttämään (ks. myös luku 5.2.4).

5.1.3 Yhteenveto

Suurin osa suomen kielen uusista yhdyssanoista sisältää 2 leksikaalista osaa. Myös kolmi- ja neliosaisia yhdyssanoja syntyy jonkin verran, mutta sitä pidemmät tapaukset ovat hyvin harvinaisia.

Suomen kielen uusien yhdyssanojen keskipituus on noin 13 merkkiä, mutta eniten aineistossa on 11 merkkiä sisältäviä sanoja. Lyhimmät tapaukset sisältävät 5 merkkiä ja pisin 33 merkkiä. Produktiivisuus kääntyy laskuun 12 merkin kohdalla,

jossa kulkee myös ihmisen tarkan näön alueen raja. Alle 10 merkkiä sisältäviä yhdyssanoja syntyy vain vähän, sillä uusien yhdyssanojen osina käytettävät perussanat ja johdokset ovat suomen kielessä pitkiä.

Periaatteessa suomen kielen yhdyssanat voivat sisältää loputtomasti sekä leksikaalisia osia että merkkejä, mutta monet sellaisetkin sanat, jotka kielenkäyttäjät pystyisivät vielä ymmärtämään ja hyväksymään kielenmukaisiksi, ovat usein liian epäkäytännöllisiä todelliseen kielenkäyttöön: pitkät yhdyssanat voivat olla vaikeita ymmärtää ja niiden käyttö on kömpelöä. Myös kielenhuollon ohjeet siitä, että tulokinnanvaraisissa tapauksissa ilmauksen osat tulisi kirjoittaa mieluummin erikseen kuin yhteen, saattavat lisätä lyhyiden yhdyssanojen suosiota ainakin niissä tapauksissa, joissa yhdyssanan osista yksi tai useampia on muussa sijamuodossa kuin nominatiivissa. Vaikka pitkien yhdyssanojen käytettävyys on huono, ne ovat kuitenkin tarpeen esimerkiksi käsitejärjestelmien tarkan kuvaamisen vuoksi.

Suomen kielessä pyritään välttämään yhdyssanan alkuosan morfologista kompleksisuutta, mikä heijastuu myös uusien yhdyssanojen osien pituuteen: uusien yhdyssanojen alkuosat ovat tilastollisesti merkitsevästi lyhyempiä kuin jälkiosat. Uusien yhdyssanojen ensimmäisen osan mediaanipituus on 5 ja viimeisten osien 6 merkkiä.

5.2 Sanaluokka

Tässä luvussa tarkastelen, kuinka paljon eri sanaluokkiin muodostetaan uusia yhdyssanoja ja mitä sanaluokkia uusien yhdyssanojen osat edustavat. Tämän luvun ensimmäisessä alaluvussa (5.2.1) esittelen käyttämäni sanaluokkajaon. Toisessa alaluvussa (5.2.2) käsittelen aineistoni kokonaisten yhdyssanojen sanaluokkaa, kolmannessa alaluvussa (5.2.3) yhdyssanojen jälkiosien sanaluokkaa ja neljännessä alaluvussa (5.2.4) yhdyssanojen alkuosien sanaluokkaa.

5.2.1 Tutkimuksessa käytetty sanaluokkajako

Tässä tutkimuksessa käytetyn sanaluokkajaon pohjana on perinteinen jaottelu kolmeen pääsanaluokkaan eli a) sijassa ja luvussa taipuviin nomineihin, b) persoonassa sekä aika- ja tapaluokissa taipuviin verbeihin sekä c) taipumattomiin ja vajaasti taipuviin sanoihin. Nominatit jakautuvat lisäksi substantiiveihin, adjektiiveihin, pronomineihin ja numeraaleihin. Taipumattomat ja vajaasti taipumattomat sanaluokat jaetaan adverbeihin, adpositioihin ja partikkeleihin (rajankäynnistä nomini- ja verbimuotojen kanssa ks. ISK § 438, 681–686; taipumattomien ja vajaasti taipuvien sanojen jaottelusta tarkemmin esim. ISK § 646, 687, 792–796, 863; aiemmin käytetyistä jaotteluista esim. Ikola 1968: 169–173). Taivutuskriteerin lisäksi olen tarvittaessa käyttänyt apuna tietoa sanan semantiikasta ja syntaktisesta käyttäytymisestä, ja esimerkiksi taipumattomat adjektiivit olen laskenut adjektiivien ryhmään (ks. myös ISK § 610).

Sanojen tarkka merkitys ja kulloinenkin sanaluokka määräytyvät viime kädessä vasta kontekstissa, mutta sanoja on perinteisesti jaettu sanaluokkiin myös sen mukaan, missä sanaluokissa ne esiintyvät tyypillisimmin tai voivat potentiaalisesti esiintyä. Tämän tutkimuksen aineistotietokannassa jokainen hakusana esiintyy vain kerran, mutta monien hakusanojen esiintymiskonteksteista on useita poimintoja. Sama sana voi edustaa eri poiminnoissa eri sanaluokkaa. Tämän vuoksi tässä tutkimuksessa esitetyt sanaluokkalaskelmat perustuvat ensisijaisesti sanojen sanaluokkapotentiaaliin. Huomioin toki myös kontekstit potentiaalin määrittelyssä.

Yhdyssanan alkuosana toimiessaan sana menettää suuren osan kieliopillisista ja syntaktisista ominaisuuksistaan (Spencer 2003: 215–216), ja kontekstista ja tulkinasta riippuen alkuosan sanaluokka voidaan määrittellä eri tavoin. Tässä tutkimuksessa perustan myös alkuosan sanaluokan määrittelyn kyseisen sanan sanaluokkapotentiaaliin (sellaisena, joka se on sanan esiintyessä itsenäisenä).

Suomen kielessä sanaluokkapotentiaalin määrittelyä hankaloittaa suomelle ja muille suomalais-ugrilaisille kielille ominainen sanaluokkajaon häilyvyys. Ongelmallisia tapauksia ovat erityisesti infiniittiset verbimuodot ja partisiipit (ISK § 490, 632). Myöskään raja substantiivien ja adjektiivien välillä ei ole kaikissa tapauksissa selvä (ISK § 625–627). Mikäli yhdyssanan osana olevalla partisiipilla on leksikkoon vakiintunut substantiivinen tai adjektiivinen merkitys ja sitä on yhdyssanassa myös käytetty tässä merkityksessä, olen luokitellut osan (ja jälkiosien tapauksissa samalla koko muodosteen) tämän leksikaalistuneen merkityksen mukaisesti (esim. *kaari-käytävä* on luokiteltu substantiiviksi, *digisopiva* adjektiiviksi). Muussa tapauksessa olen luokitellut partisiipit omaan kategoriaansa: partisiipeiksi (esim. *muuntokoulutettu*, *ummistettusilmäinen*). Joissakin tapauksissa olen käyttänyt leksikaalistumista myös adjektiivien ja substantiivien erottelukeinona (esim. *vanha*-sanan vertailumuodot on tulkittava substantiiveiksi senkaltaisissa sanoissa kuin *bumerangivanhempi* ja *sukellusvanhin* vrt. esim. *ottovanhempi* ja *kylänvanhin*).

Koska joidenkin suomen kielen sanojen sanaluokkapotentiaaliin kuuluu, että niitä voidaan käyttää useissa sanaluokissa, olen perinteisten sanaluokkien lisäksi käyttänyt luokittelussa useaa sanaluokkaa edustavien sanojen ryhmää. Kokonaisten yhdyssanojen joukossa kaikilla useaa sanaluokkaa edustavilla sanoilla on käytännössä substantiivin ja adjektiivin sanaluokkapotentiaali (tyypillisiä tapauksia ovat esimerkiksi värinimitykset sekä monet ihmiseen viittaavat sanat, esim. *retronruskea*, *han-kiinalainen*). Useaa sanaluokkaa edustavien yhdyssanan alkuosien kohdalla kirjo on sen sijaan laajempi. Alkuosalla voi olla ainakin substantiivin ja adjektiivin (esim. *hämäräkytkin*), substantiivin ja verbin (esim. *hyökylurma*) sekä pronominin, adverbien ja partikkelin (esim. *itsepöhö* 'itserakkaus, röyhkeys, arroganssi tms.') sanaluokkapotentiaali.

5.2.2 Koko muodosteen sanaluokka

Tässä luvussa käsittelen suomen kielen uusien yhdyssanojen sanaluokkajakaumaa. Luvussa tarkasteltavat yhdyssanarakenteet eli uusien yhdyssanojen sanaluokat eivät ole kielenkäytössä tyypillisesti keskenään vaihtoehtoisia: eri sanaluokat eivät

kilpaile yhdyssanojen muodostuksessa keskenään toisin kuin esimerkiksi erilaiset alkuosan muodot tai alkuosan sanaluokkakin voivat tehdä (esim. sanoilla *tyvikohottaja* [TT] ja *tyvenkohottaja* [TT] tai *uimaputkilo* [TT] ja *uintiputkilo* [TT] voidaan viitata samaan asiaan mutta sanoilla *tyvikohottaja* ja *tyvikohottaa* ei voida). Kysymys suomen kieleen syntyvien uusien yhdyssanojen sanaluokasta onkin olennainen ennemmin kielen sananmuodostustendenssien kuin eri yhdyssanarakenteiden keskinäisen kilpailun kannalta (vaikkapa yhdyssubstantiivien suosio sananmuodostuksessa voi vaikuttaa esim. uusien johtamalla muodostettujen substantiivien määrään, mutta ei todennäköisesti yhdysverbien määrään). Näin ollen tässä luvussa prosenttiosuuksia olennaisempia ovat absoluuttiset luvut. Vertaamalla tutkimusaineistoa toisten tutkimusten tuloksiin voidaan kuitenkin tehdä päätelmiä myös tiettyä sanaluokkaa edustavien yhdyssanojen suosiosta, joten myös tieto prosenttiosuuksista on tarpeen.

Tässä luvussa on käytetty osa-aineistoa 1, joka sisältää yhden kirjoitusasun kaikista aineiston yhdyssanoista ($n = 28\ 013$; osa-aineistot on esitelty tarkemmin luvussa 4.3). Tiedot aineiston yhdyssanojen sanaluokkajakaumasta on koottu taulukkoon 4.

TAULUKKO 4. Tutkimusaineiston yhdyssanojen sanaluokkajakauma ($n = 28\ 013$)

Sanaluokka	Esimerkki	Osuus	Määrä
Substantiivi	<i>heimomainonta, silikonikinnas</i>	92,8 %	26 004
Adjektiivi	<i>kulutuskriittinen, kännykkävapaa</i>	4,1 %	1 139
Verbi	<i>luontomatkailla, neulahuovuttaa</i>	1,2 %	328
Partisiippi	<i>hetitappava, lankavärjätty</i>	0,9 %	265
Substantiivi-adjektiivi	<i>helsinginharmaa, lihassairas</i>	0,7 %	203
Taipumaton tai vaillinaisesti taipuva sanaluokka	<i>kertalinttuulla, puoliseluisesti</i>	0,3 %	73
Pronomini	<i>ei-mikään</i>	0,004 %	1
Kaikki sanaluokat		~ 100 %	28 013

Substantiivit ovat uuden yhdyssanaston ylivoimaisesti yleisin sanaluokka. Niitä on aineistossa 26 004 kpl (92,8 %). Seuraavaksi eniten on adjektiiveja (1 139 kpl, 4,1 % aineistosta) ja verbejä (328 kpl, 1,2 % aineistosta). Kaikkiin muihin sanaluokkiin kuuluvia tapauksia on kutakin alle 1 % aineistosta.

Substantiivien ylivalta on todettu myös suomen kielen vakiintunutta yhdyssanastoa tarkastelevissa tutkimuksissa (Saukkonen 1973; Jussi Niemi 2009). Saukkosen (1973: 388) mukaan NS:n yhdyssanoista noin 89 % on substantiiveja, noin 10 % adjektiiveja, noin 0,7 % taipumattomiin tai vajaan taipuviin sanaluokkiin

kuuluvia (Saukkosella adverbija tai partikkeleita) ja noin 0,3 % verbejä. Jussi Niemi (2009: 251) ei mainitse CD-PS:n kaikkien yhdyssanojen määrää (taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvat sanat mukaan luettuna), mutta pelkästään substantiivien, adjektiivien ja verbien joukossa eri sanaluokkien jakauma on hänen aineistossaan 89,7 % substantiiveja, 9,8 % adjektiiveja ja 0,6 % verbejä⁶³. Saukkosen (1973) ja Niemen (2009) suomen yleiskieltä kuvaavista sanakirjoista saamat tulokset ovat siis keskenään hyvin samansuuntaisia.

Uuden yhdyssanaston sanaluokkajakauma (ks. taulukko 7 edellä) eroaa tilastollisesti merkitsevästi⁶⁴ sekä NS:n (Saukkonen 1973, Fisherin tarkan testin $n = 2\ 000$, $p < 0,001^{***}$) että CD-PS:n (Jussi Niemi 2009, Pearsonin khin neliö -riippumattomuustestin $\chi^2 = 22,3$, $n = 2\ 000$, $df = 2$, $p < 0,001^{***}$) sanaluokkajakaumasta. Sovitettujen standardoitujen jäännösten avulla tehty jatkotarkastelu paljastaa, että tilastollisesti merkitsevä ero on seurausta adjektiivien vähäisyydestä uudessa yhdys-sanastossa (adjektiivien sovitettu standardoitu jäännös vertailussa NS:aan = 3,7 ja vertailussa CD-PS:aan = 3,2). Muiden sanaluokkien osuuksissa ei ole tilastollisesti merkitseviä eroja.

Varsinkin tilapäisluonteiset *-inen*-loppuiset adjektiivit voivat helposti jäädä sanakirjojen ja jopa uudissanatietokantojen ulkopuolelle vahvan kontekstisidon-naisuutensa vuoksi (esim. 10 ja 11).

- (10) Yrttivalmisteista mukula-artisokka- ja merileväuutteet voivat vilkastuttaa aineenvaihduntaa ja siten tukea laihdutusta. Mukula-artisokka sisältää *haimaystävällistä* inuliinia, joka auttaa vatsan hyvinvointia ja hillitsee ruokahalua. [TT <- ET-lehti, 1/2002, s. 73]
- (11) Miksi ihmeessä tämä pariskunta on kuluvalle viikolla komeillut joka jumalan lööpissä? Iskelmälaulaja ja kotiseutu-kirjailija – eikö näinä *missintäyteisinä* aikoina raflaavampaa uutista löydy? [TT <- Metro, 13.2.2004, s. 6]

Tämä ei kuitenkaan selitä sitä, että yhdysadjektiivit ovat varsin sekalaista uutta yhdyssanastoa sisältävässä tietokannassa harvinaisempia kuin sanakirjoissa. Yhdysadjektiivien suhteellinen osuus näyttää siis olevan pienemässä. Tämä ei kuitenkaan välttämättä tarkoita, että uusia yhdysadjektiiveja syntyisi kieleen vähemmän kuin ennen; ehkä uusia yhdysubstantiiveja ja -verbejä vain syntyy aiempaa nopeammalla tahdilla. Adjektiivien kohdalla on todennäköisesti myös muita sanaluokkia enemmän mahdollisuuksia käyttää yhdyssanoja korvaavia sanaliittoja, sillä varsinkin genetiivialut ovat tyypillisiä nimenomaan yhdysadjektiiveille (ks. myös luku 7.2.1). Muuta syytä yhdysadjektiivien pienelle osuudelle uusien yhdyssanojen joukossa on vaikea löytää, sillä suomen kieli ei todennäköisesti ole yhtäkkiä alkanut

63 Osuudet laskettu Niemen raportoimista luvuista tätä tutkimusta varten. Niemen mukaan CD-PS:ssa on kaikkiaan 51 356 substantiivia, 5607 adjektiivia ja 317 verbiä.

64 Testauksessa ovat tämän tutkimuksen aineistosta mukana vain Saukkosen (1973) ja Jussi Niemen (2009) käyttämiin sanaluokkiin kuuluvat yhdyssanat. Ylipäänsä olen kaikissa vertailuissa rajannut omaa aineistoani mahdollisimman tarkoin vertailukohteen mukaan (esim. Niemen aineistonrajauksesta ks. Niemi 2009: 250–251).

karttaa kuvailevuutta tai hankkia nimenomaan kuvailevaa sanastoa esimerkiksi lainaamalla.

Kielenhuolto on jo pitkään ohjeistanut välttämään monia (alkujaan lainaperäisinä) muoti-ilmauksina pidettyjä yhdysadjektiiveja ja yhdyssanojen adjektiiviloppuja (esim. Saarimaa 1971 [1947]: 74–78; Vesikansa 1989: 241–250; Palander 2002). Niin sanottuja muotiloppuja käytetään kuitenkin yhdysadjektiivien muodostuksessa edelleen paljon, eikä niiden välttäminen siis selitä yhdysadjektiivien pientä osuutta aineistossa. Aineistossa toistuvat usein esimerkiksi seuraavat kielenhuoltajien joko nykyään tai aiemmin vältettäväksi suosittamat yhdysadjektiivien loppuosat: *-herkkä* (17 kpl), *-ystävällinen* (14 kpl), *-vapaa* (13 kpl), *-varma* (7 kpl), *-intensiivinen* (6 kpl), *-rikas* (6 kpl), *-tarkka* (5 kpl), *-tehokas/tehoton* (yht. 5 kpl), *-köyhä* (4 kpl), *-kypsä* (3 kpl), *-tiivis* (3 kpl), *-kylmä* (1 kpl) ja *-täyteinen* (1 kpl) esiintyvät yhteensä 85:ssä aineistoni sanassa eli 7,6 %:ssa⁶⁵ kaikista aineistoni yhdysadjektiiveista (kielenhuoltajien antamista sanakohtaisista ohjeistuksista ks. tarkemmin esim. Vesikansa 1989: 241–250; Palander 2002: 15–17; eri adjektiiviloppujen yleisyydestä myös vakiintunutta sanastoa sisältävässä lehtikielessä M. Karvinen 1999).

Vanhastaan muotilopuiksi mainittujen sanojen lisäksi aineistossani on paljon myös uusia muotimyönteisiä tarjokkaita. Osaa yhdysadjektiivien muotilopuista on moitittu esimerkiksi siitä, että ne jättävät koko muodosteen merkityksen hämäräksi (esim. Palander 2002: 17; ks. myös Iisa–Oittinen–Piehl 2002: 216). Uusi merkitykseltään hämärä yhdysadjektiivien muotiloppu on *-lähtöinen*, joka esiintyy esimerkiksi sellaisissa sanoissa kuin *homelähtöinen*, *prosessilähtöinen* ja *tuotantolähtöinen*. Yhteensä loppuosa *-lähtöinen* esiintyy aineistoni yhdysadjektiiveissa 16 kertaa. Muita muoti-ilmauksen kaltaisia ja usein aineistossa toistuvia loppuosia ovat esimerkiksi *-vetoinen* (13 kpl), *-peräinen* (12 kpl), *-hakuinen* (10 kpl), *-perustainen* ja *-perusteinen* (yhteensä 9 kpl), *-pihi* (8 kpl), *-myönteinen* (7 kpl) ja *-neutraali* (6 kpl).

Esimerkiksi Marjatta Palanderin (2002, ks. myös Vesikansa 1989: 241–250; M. Karvinen 1999) mukaan kielenhuoltajien ohjeet ovat auttaneet hävittämään joitakin kartettaviksi suositettuja muoti-ilmauksia kielestä. Kuten oma aineistoni osoittaa ja Palanderkin toteaa, yhdysadjektiiviongelma ei kuitenkaan ole kadonnut minnekään, sillä monet kartettavina pidettävät adjektiiviloput ovat sananmuodostuksessa edelleen tiheässä käytössä. Kielenhuoltoakaan ei siis todennäköisesti voida syyttää yhdysadjektiivien pienestä osuudesta uudessa yhdyssanastossa.

Uuden yhdyssanaston ja vakiintunutta sanastoa sisältävän sanakirja-aineiston välillä ei ole tilastollisesti merkitsevää eroa yhdysverbien määrässä. Uudessa yhdyssanastossa tyypillisiä ovat todennäköisesti erikoiskielistä alkunsa saaneet termimäiset ilmaukset (esim. *kuvaikutuskata*, *massiivilistoittaa*). Toinen suuri ryhmä ovat tyyliään keveät tai leikkisät ilmaukset (esim. *juustohöylätä* 'tehdä leikkauksia', *lemmenleikkiä*). Kuten esiin nostamistani esimerkeistäkin näkyy, monet uusista yhdysverbeistä on muodostettu johtamalla aiemmista yhdyssanoista. Johtaminen

65 Laskelmasta on jätetty pois senkaltaiset kotoisia rakenteita noudattavat (joskin toisinaan osin vieraista aineksista koostuvat) ilmaukset kuin *antikaunis*, *ultrapuhdas*, *uusvapaa* 'uudelleen sinkukuksi tullut' ja *vahvanherkkä*.

on mainittu yhdysverbien tyypillisenä muodostuskeinona myös monissa aiemmissä tutkimuksissa (esim. Vesikansa 1989: 256–258; Tuomi 1995: 154; Levonen 2006). Käsittelen eri sanaluokkiin kuuluvien yhdyssanojen muodostustapoja tarkemmin luvussa 6.1.1.

Monet tutkimusaineiston yhdysverbeistä ovat vahvasti kontekstisidonnaisia, eivätkä ne todennäköisesti koskaan tule vakiintumaan kieleen. Kuitenkin myös tilapäismuodosteiden rakenne kertoo siitä, millä keinoin kielenkäyttäjät katsovat parhaaksi tiivistää informaatiota yhdyssanoihin. Vahva kontekstisidonnaisuus on tyypillistä myös yhdyspartisiipeille (esim. *naisvärvätty* 'naispuolinen sotilaaksi värvätty', *videotykitetty*). Kuitenkin myös niiden joukossa on useita jo tämän tutkimuksen tekohetkeen mennessä ainakin jossain määrin adjektiiviksi tai substantiiviksi vakiintuneita ilmauksia (esim. *ykkössiioitettu*, *ostovoimakorjattu*, *homealtistunut*, *biohajoava*, *tohtorikoulutettava*).

Myöskään taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvien sanojen määrissä ei ole tilastollisesti merkitseviä eroja tämän tutkimuksen aineiston sekä NS:n ja CD-PS:n välillä. Olennainen ero aineistoissa on kuitenkin se, että toisin kuin NS:ssa (Saukkonen 1973: 338), omassa aineistossani ei ole lainkaan adpositioita tai partikkeleita, vaan kaikki taipumattomien sanojen ryhmään kuuluvat sanat ovat adverbejä. Tämän johtuu siitä, että adpositioiden ja partikkelien luokkiin syntyy uusia sanoja ylipäänsä harvoin ja että adpositiot ja partikkelit ovat harvoin yhdyssanoja.

Kuten osa yhdysadjektiiveista, myös osa adverbeistä on erilaisissa aineistoissa eri asemassa. Mikkolan (1971) tutkimassa kaunokirjallisuuden kielen otoksessa yhdysadverbejä on peräti 6,3 % koko aineiston yhdyssaneista (taipumattomiin tai vajaasti taipuviin sanaluokkiin kuuluvia yhdyssanoja yhteensä 7,2 %), kun tämän tutkimuksen aineistossa adverbejä on vain 0,3 %. Adverbeistä varsinkin *-sti*-adverbit⁶⁶ saattavat olla sanastotietokannoissa ja myös sanakirjoissa aliedustettuina siihen nähden, kuinka paljon niitä kielessä käytetään, sillä *-sti*-adverbit ovat usein seurausta niin sanotusta syntaktisesta johtamisesta (syntaktisesta johtamisesta ks. tarkemmin esim. Koski 1978: 103). Tämän tutkimuksen aineiston adverbeistä *-sti*-adverbejä on 39,7 % (29/73 kpl). Valtaosa aineiston *-sti*-adverbeistä on muodostettu liittosanasta tai aiemmasta yhdyssanasta johtamalla (esim. *ennätyssakeasti*, *syväpuhtaasti*; yhdyssanojen muodostustavoista ks. tarkemmin luvut 2.2.3 ja 6.1).

Uusia yhdysadverbejä muodostetaan myös yhdistämällä. Muodostuksessa toistuu joitakin selviä kiinteän leksikaalisen osan sisältäviä yhdyssanamuotteja (esim. [*puoli* + ADV] > *puolisattumalta*, *puolivahingossa*, *puolivakavissaan*; [SUBST + mielessä] > *konseptimielessä*, *showmielessä*). Lisäksi yhdistämisessä käytetään hyödyksi tietoa siitä, että yhdyssanan osien suhde voi olla käytännössä mikä tahansa. Esimerkiksi yhdysadverbissa *pyyhemunasillaan* (esim. 12) muodosteen molemmat osat tuovat lisänsä koko muodosteen merkitykseen '(miehestä:) olla alasti kuitenkin

66 *-sti*-muotojen (ja niitä muissa kielissä vastaavien rakenteiden) asemasta kieliopissa ei olla yksimielisiä, ja niitä on toisinaan pidetty myös adjektiivin taiputusmuotoina (esim. Tuomikoski 1973). Tässä tutkimuksessa *-sti*-muodot on kuitenkin katsottu adverbeiksi (ks. myös ISK § 373).

pyyhe verhonaan⁶⁷, mutta osien välinen merkitys yhteys ei ole läpinäkyvä tai selvä ilman kulttuurikontekstin tuntemusta.

- (12) Periltä löytyy viisitoista höyryävää miestä *pyyhemunasillaan* olutpullojen ympäröiminä. [TT <- HS Nyt 11, 15.–21.3.2002, s. 47]

Vaikka uusia yhdyspronomeja ja -numeraaleja ei näytä käytännössä syntyvän, myös näiden sanaluokkien laitamille voi toisinaan poikkeuksellisesti ilmaantua uusia yhdyssanoja. Omassa aineistossani ainoa uusi pronominiksi luokiteltavissa oleva yhdyssana on *ei-mikään*⁶⁷ (esim. 13).

- (13) Keskellä *ei-mitään* Alma-Atassa Stenforsille valkeni, miltä tuntuu olla syrjässä. [TT <- Talouselämä 28, 31.8.2001, Seija Holtari, s. 74]

Ilmauksessa kieltoverbin ja kvanttoripronominin liitto on tiivistynyt ja kiteytynyt tietynlaista ympäristöä merkitseväksi yhdyssanaksi. Vaikka yhdyssanalla on jonkinlainen tilanteesta riippumaton merkitys, siltä kuitenkin edelleen puuttuu substantiiveille tyypillinen tarkka luokittava merkitys (ISK § 713). Semantiikaltaan sana jääkin substantiivien ja pronomien välimaastoon. Pronominiksi luokittelua tukee se, että ilmauksen loppuosa on säilyttänyt pronominille kuuluvan taivutuksensa (*keskellä ei-mitään* mutta ei **ei-mikäättä*).

Ainoa aineistosta kokonaan puuttuva nominaalinen sanaluokka on numeraalit. Yhdysnumeraaleja voidaan tietysti muodostaa loputtomasti. Koska kaikki perusluvut ovat jo tunnettuja, niiden nimityksiä ei kuitenkaan voida pitää uudissanoina, vaikka kukaan ei sattumalta olisi aiemmin puhunut juuri kyseisestä (suuresta) luvusta. Yhdysnumeraaleja ei siksi kerätä uudissanastotietokantoihin (poikkeuksena numeraalien substantiivijohdokset, esim. *ysiysi* ’99-laadun bensiini’, jotka on tässä tutkimuksessa luettu substantiivien ryhmään).

5.2.3 Jälkiosan sanaluokka

Tässä luvussa esitän muutamia tarkennuksia perinteiseen fennistiseen käsitykseen yhdyssanan jälkiosan sanaluokan ja koko muodosteen sanaluokan yhteydestä. Tämä luku ei sisällä tilastollista analyysia tutkimusaineistostani, sillä suomen kielessä yhdyssanan jälkiosan sanaluokka on tyypillisesti sama kuin jo edellä käsittelemieni kokonaisten muodosteidenkin (ISK § 404, ks. myös edellinen luku 5.2.2). Käytän kuitenkin luvussa aineiston sanoja esimerkkitaipauksina.

Jälkiosa on suomen kielen yhdyssanoissa kieliopillisesti hallitseva jäsen. Se määrää yhdyssanan syntaktisen ja morfologisen käyttäytymisen, joten koko yhdyssanan sanaluokka on sama kuin sen jälkiosan sanaluokka. Määrittämissä yhdyssanoissa

67 Vrt. englannin kieltävät pronominit ja adverbit (esim. *nobody*, *nothing* ja *nowhere*). Yhteen kirjoitettuja *ei*-alkuisia pronomeja esiintyy myös Agricolan kielessä (esim. *eikengen*, *eikellenge*). Poimimiskontekstissaan sana esiintyy muodossa *ei-mitään*, mutta NSST:ssa sen lemmattuna hakusanamuotona on *ei-mikään*. Myös muut NSST:n Hakusanat on lemmattu NSST:ssa perusmuotoihinsa. Tässä tarkastelun kohteena olevassa sanassa lemmaaminen on sikäli kyseenalaista, että kyse on ilmeisesti vain tietyssä muodossa esiintyvistä käännoislainasta.

jälkiosa on yleensä myös semanttisesti hallitseva jäsen, kun taas summayhdyssanoissa osat ovat semanttisesti samanarvoisia. (ISK § 404, 432.) Tarkasteltaessa yhdyssanan sanaluokan määräytymistä on huomattava erottaa yhdyssosan semanttinen ja kieliopillinen pääjäsenyys toisistaan (pääjäsenyyden käsitteestä tarkemmin esim. Bloomfield 1933: 35; Williams 1981; Zwicky 1985; Croft 1996; Štekauer 2001). Tärkeää on myös eronteko sanan potentiaalisen sanaluokan ja kulloisessakin kontekstissa aktuaalistuvan sanaluokan välillä. Koska yhdyssanan jälkiosa on suomen kielessä aina⁶⁸ muodosteen kieliopillisesti hallitseva jäsen, yhdyssanan kontekstisanaluokka on suomessa aina sama, jossa sen jälkiosa kulloisessakin kontekstissa esiintyy (tosin kaikissa konteksteissa sanaluokkaa ei voida yksiselitteisesti määrittää, ks. esim. ISK § 625–626, 632).

Yhdyssanan osien yhdessä luoma sanan sisäinen mikrokonteksti voi olennaisesti muuttaa jälkiosan sanaluokkapotentiaalia. Esimerkiksi yhdyssanan mikrokontekstista käy vaikkapa ravikielen sana *hiittisuora*, jossa loppuosa *suuraa* määrittävä *hiitti* 'kovavauhtinen harjoitusajo' osoittaa loppuosan esiintyvän tässä substantiivisessa merkityksessä 'radan suora osa'. Koska määrittäsyhdyssanoissa alkuosa määrittää jälkiosaa ja siten myös usein osoittaa sen kulloisenkin merkityksen, yhdyssanan potentiaalinen sanaluokka määräytyy oikeastaan ennemmin sen osien yhdessä luomasta mikrokontekstista ja koko muodosteen merkityksestä kuin pelkästä jälkiosan potentiaalisesta sanaluokasta. Voidaankin esimerkiksi olettaa, että monet sekä adjektiivin että substantiivin sanaluokkapotentiaalin omaavat sanat esiintyvät useammin substantiiveina toimiessaan yhdyssanan jälkiosana kuin itsenäisenä lekseeminä. Esimerkiksi sanat *kylmä* ja *hyvä* ovat potentiaaliltaan ensisijaisesti adjektiiveja, mutta esimerkeissä 14 ja 15 ne esiintyvät yhdyssanan substantiivisina jälkiosina. Sekä *kylmä* että *hyvä* voivat toki esiintyä monenlaisissa konteksteissa substantiiveina myös yksinään, mutta yhdyssanoissa *kaukokylmä* ja *lisähyvä* ne osoittaa substantiiviksi jo yhdyssanan merkitys ilman tietoa kontekstista. Yhdyssanan mikrokonteksti siis konkretisoi abstraktin adjektiivin substantiiviksi ja osoittaa jälkiosan potentiaalisista sanaluokista mahdolliseksi vain toisen.

- (14) Koska kaukolämmölle ei ole niin paljon kulutusta, – ylijäämälämpöä joudutaan lauhduttamaan meriveteen. Ratkaisuksi onkin kehittymässä uudentyyppinen hyötymuoto: *kaukokylmä*. – Ylijäämälämpöä pystytään jo hyödyntämään niin, että teknisen prosessin avulla siitä tehdäänkin *kaukokylmää*. [TT <- Helen – Helsingin Energian asiakaslehti, 2/2005, s. 13]
- (15) Yksin jo tuollaisen prässin kestäminen ”kaalin” hajoamatta on kultainen suoritus. 90 metrin keihäskaari on sitten mukavaa *lisähyvää*. [TT <- Uutispäivä Demari, 4.9.2007, Rolf Bamberg, s. 4]

Tyypillisiä esimerkkejä tällaisista sanoista ovat myös useat *-minä*-loppuiset yhdyssanat, joissa yhdyssanan luoma mikrokonteksti nostaa esiin *minä*-sanan

68 Tarkalleen ottaen numeraalit 11–19 ovat poikkeus.

harvinaiset substantiiviset merkitykset. Substantiivius näkyy myös taivutuksessa sekä *minä*-sanana ja puheaktipersoonan deiktisen suhteen puuttumisena (esim. 16 ja 17).

- (16) Vaalien jälkeen entisestä oppositiopuolueesta voi kuoriutua uusi hallituspuolue, joka muistuttaa *oppositiominäänsä* kuin toukka perhosta. [TT <- SK, 8.3.2002, s. 83]
- (17) Katajavuori on häivyttänyt *runominän* pääosin taustalle, osaksi arjen tarinaa. [TT <- HS, 1.12.2001, s. B2]

Yhtenä yhdyssanuuden kriteerinä on pidetty jälkiosan semanttista pääjäsenyyttä (Mäkisalo 2000b: 577). Englannin kieleen keskittyvässä tutkimuksessa on sovellettu niin sanottua RHR-sääntöä (*right-hand head rule*), jonka mukaan morfologisesti kompleksisen sanan semanttinen pääjäsen on aina jäsenistä äärimmäisenä oikealla oleva (ks. esim. Williams 1981). Sääntö pätee pääsääntöisesti myös muihin germaanisiin kieliin, mutta esimerkiksi romaanisissa kielissä yhdyssanan semanttinen pääjäsen on tyypillisesti muodosteen ensimmäinen osa (Scalise–Guevara 2005: 168). Poikkeuksettomana sääntöä ei voida pitää edes englannissa (englannin ja muiden kielten osalta Booij 2007: 76; Scalise–Guevara: mp.).

Myöskään suomessa yhdyssanan semanttinen pääjäsen ei välttämättä aina ole muodosteen jälkijäsen. Summayhdyssanoissa on kaksi semanttista pääjäsentä, sillä yhdysosat ovat niissä semanttisesti tasa-arvoisia (esim. *kahvila-ravintola* on sekä kahvila että ravintola; ISK § 432). Määrittämissuhteidenkin kohdalla esimerkiksi koostumussuhteisia käsitejärjestelmiä heijastelevien termien semanttiseksi pääjäseneksi voidaan käsittää myös muodosteen alkuosa. Koostumussuhteisen käsitejärjestelmän näkökulmasta esimerkiksi lehden osiin viittaavissa sanoissa *lehtilapa* ja *lehtiruoti* jälkiosa määrittää semanttisena pääjäsenenä toimivaa alkuosaa. Sama termi voi myös olla alku- tai jälkimääritteinen riippuen siitä, mihin käsitejärjestelmään se kulloinkin kuuluu. (Pitkänen 2008: 202–203, esimerkit Pitkäsen.) Alkuosa on semanttisesti hallitseva myös sentyyppisissä yhdyssanoissa kuin *serkkutyttö*, *rovastivainaja*, *lehmäkanttura* (Vesikansa 1989: 214, 223–224). Monet tämääntyyppisistä sanoista esiintyvät myös genetiivialkuisina (*lehmänkanttura*, *miehenpuolikas*), ja monet niistä voidaan kirjoittaa joko yhdyssanaksi tai sanaliitoksi (*äitikulta* ~ *äiti kulta*). Myös omassa aineistossani on jotakin tämääntyyppisiä tapauksia. Esimerkissä 18 loppuosa *-jeesus* on muodosteessa vailla omaa semanttista sisältöä oleva elementti, jonka tehtävä on lisätä sanottavan painokkuutta: 'ei yhtäkään penniä, ei pennin()pyörylää'.

- (18) Yleensä palkkio tulee, jos asetetut talouteen liittyvät tulostavoitteet tai muut päämäärät saavutetaan. Siitä, että toimihenkilö yrittää pysyä terveenä, ei kukaan tietävästi ole hoksannut maksaa ikinä missään *penninjeesusta*. [TT <- Kauppalehti Optio, 19.10.2000, s. 13]

Yhdyssanan semanttisen pääjäsenen ja koko yhdyssanan sanaluokan välillä on vahva yhteys. Tämä näkyy parhaiten sellaisista yhdyssanoista, joissa jälkiosa voisi potentiaalisesti edustaa eri sanaluokkaa kuin koko muodoste. Esimerkiksi sanan

autonrämä semanttinen pääjäsen on sen substantiivinen alkuosa *auto*. Alkuosaa määrittävää jälkiosaa *rämä* voidaan yksinään käyttää sekä substantiivina että adjektiivina. Yhdyssana *autonrämä* on kuitenkin substantiivi. Semanttinen pääjäsen ei kuitenkaan aina määrää koko muodosteen sanaluokkaa. Tämä huomataan helpoiten eksosentrisistä yhdyssanoista, joissa ei yleensä katsota olevan lainkaan semanttista pääjäsentä, sillä ne eivät ole edusosansa alakäsitteitä (esim. lepakkoon viittaava *nahkasiipi* ei ole sen enempää nahka kuin siipikään). Tällaisissa yhdyssanoissa kokonaisuus voi toisinaan edustaa sellaista sanaluokkaa, joka ei tunnu kuuluvan sen kummankaan osan sanaluokkapotentiaaliin (esim. 19).

- (19) Arkipuku muuttuu juhlavaksi pienellä vaivalla: *sävysävy* liina taskuun ja siinä se. [TT <- City Styleguide, marraskuu/2005, s. 34]

Esimerkissä 19 on kahdesta *sävy*-substantiivista koostuva toistoyhdyssana, joka toimii kontekstissaan adjektiivina. *Sävysävy*-adjektiivin muodostustapa on sikäli poikkeuksellinen, että se on johdettu takaperoisesti sanaliitosta *sävy sävyyn* 'värisävylytään toisiinsa sointuvasti'. Myös osan kautta kokonaisuuteen viittaavia bahuvriihiyhdyssanoja, yhtä eksosentristen yhdyssanojen alalajia, on vanhastaan pidetty adjektiiveina (Hämäläinen 1883: 122–124; Airila 1938: 358; Hakanen 1973: 125–126; L. Hakulinen 1979: 384–385), vaikka niiden jälkiosat ovat tyypillisesti substantiiveina esiintyviä sanoja. Myös suurin osa bahuvriihien esiintymistä on substantiivisia (Heinonen 2001: 625), mutta niitä voidaan kuitenkin käyttää myös adjektiiveina, vaikka kummallakaan osalla ei periaatteessa olisi adjektiivin sanaluokkapotentiaalia (esim. *kivikasvolla ukolla*).

5.2.4 Alkuosan sanaluokka

Tässä luvussa käsittelen suomen kielen uusien yhdyssanojen alkuosien sanaluokajakautamaa. Aihe liittyy yhdyssanatutkimuksessa tärkeänä pidettyyn kysymykseen eri sanaluokkiin kuuluvien sanojen yhdistymismahdollisuuksista (ks. tarkemmin 7.2.2). Yhdyssanan alkuosien kategorisointi sanaluokan mukaan on väistämättä josain määrin keinotekoista, sillä sanat menettävät yhdyssanan alkuosana toimiessaan suuren osan syntaktisista ja toisinaan myös semanttisista ominaisuuksistaan (Spencer 2003: 215–216; ks. myös luku 2.2.3). Koska yhdyssanan alkuosan sanaluokan tarkka määrittäminen ei ole aina mahdollista, tässä luvussa on käytetty perinteisten luokkien lisäksi myös sanaluokaltaan epäspesifien tapausten luokkaa (ks. tarkemmin jäljempää tästä luvusta, s. 120) sekä myös kokonaisten muodosteiden käsittelyssä käytössä ollutta useaa sanaluokkaa edustavien tapausten luokkaa (ks. luvut 5.2.1 ja 5.2.2).

Tässä luvussa on käytetty osa-aineistoa 3, joka sisältää vain kaksiosaisia yhdyssanoja ($n = 22\ 144$; osa-aineistot on esitelty tarkemmin luvussa 4.3). Tiedot aineiston yhdyssanojen alkuosan sanaluokajakautamasta on koottu taulukkoon 5.

TAULUKKO 5. Tutkimusaineiston yhdyssanojen alkuosan sanaluokkajakauma ($n = 22\ 144$)

Alkuosan sanaluokka	Esimerkki	Osuus	Määrä
Substantiivi	<i>hintajousto, karvaeläin</i>	79,8 %	17 680
Sanaluokaltaan epäspesifi	<i>esiteini, digionnela</i>	9,3 %	2 052
Adjektiivi	<i>kevyttakka, pikkuvappu</i>	5,6 %	1 238
Useaa sanaluokkaa edustava	<i>itsepöhö, sairasteurastettava, liukuvärjäys</i>	3,2 %	714
Taipumaton tai vajaasti taipuva sanaluokka	<i>kanssatoimija, mukaradikaali</i>	0,7 %	153
Verbi	<i>istumaannousu, syömähäiriö, kiilugeeni</i>	0,5 %	108
Pronomini	<i>minäsuuntautuneisuus, moni-identiteettisyys</i>	0,5 %	101
Numeraali	<i>nollaraja, tusinapakki</i>	0,4 %	90
Partisiippi	<i>nuokkuvakukkainen, ummistettusilmäinen</i>	0,04 %	8
Kaikki muodot		~ 100 %	22 144

Kuten taulukosta 5 nähdään, suurin osa suomen kielen uusien yhdyssanojen alkuosista on substantiiveja (79,8 %). Niiden jälkeen suurin ryhmä on sanaluokaltaan epäspesifit alkuosat (9,3 %). Seuraavaksi yleisimpiä ovat adjektiivit (5,6 %) ja useaa sanaluokkaa edustavat alkuosat (3,2 %). Muita sanaluokkia edustavia alkuosia on aineistossa alle 1 % kutakin.

Valtaosa uusien yhdyssanojen substantiivialkuosista on tavanomaisia substantiiveja, joilla on muodosteessa asiaan tai olioon viittaava substantiivinen merkitys (esim. *asiakaskortti* 'kortti, joka osoittaa (kanta)asiakkuuden', *logokonvehti* 'konvehti, jonka päällä on logo'). Osaltaan substantiivialkuosien ylivaltaa sekä uudessa että vakiintuneessa (ks. Jussi Niemi 2009: 251) yhdyssanastossa selittää myös se, että adjektiivisia ja verbaalisia merkityksiä edustaa yhdyssanoissa usein substantiivimuotoinen alkuosa. Eronteko varsinkin adjektiivisten ja substantiivisten merkitysten välillä on vaikeaa, sillä määrittelyyhdyssanoissa alkuosat ovat suhteessa edusosaansa aina sikäli adjektiivisia, että ne luonnehtivat tavalla tai toisella edusosan ominaisuuksia. Tämä näkyy selvästi esimerkiksi sellaisissa kahden substantiivin muodostamista yhdyssanoissa, joissa määriteosa ilmaisee edusosan materiaalin. Merkityksen kannalta ei olekaan selvää, onko esimerkiksi *karvaeläin*-yhdyssanan alkuosa tulkittava substantiiviksi ('eläin, jossa on karva[a]') vai adjektiivin yhdysosamuodoksi ('karvainen eläin'). Muodon kannalta alkuosa on kuitenkin selvästi substantiivi. Tässä tutkimuksessa katson tämän tyyppisistä tapauksista adjektiivin yhdysosamuodoiksi ainoastaan ne, joissa alkuosa muistuttaa kyllä muodoltaan substantiivia mutta ei vastaa mitään suomen kielen itsenäistä lekseemiä (esim. *reaalitalous* < *reaalinen talous*, ks. tästä tyyppistä tarkemmin luku 5.3.2.3).

Muodon mukaan substantiiveiksi olen laskenut myös yhdyssanan alkuosissa verbaalisia merkityksiä edustavat deverbaalit teonnimisubstantiivit (esim. *johtokaksikko* 'kaksikko, joka johtaa'). Teonnimisubstantiivi voi olla myös sellainen, että sitä ei juuri tavata itsenäisessä käytössä (esim. *jättösarvet*, *jääntisiemen*). Varsinkin

tällöin alkuosan sanaluokan valinta muistuttaa syntaktista johtamista, jossa sanan sanaluokkaa muutetaan sille syntyneiden syntaktisten vaatimusten mukaisesti (syntaktisesta johtamisesta tarkemmin esim. Koski 1978: 103). Yhdyssanojen muodostuksessa syntyvät vaatimukset sanaluokan muuttamisesta ovat sanakohtaisia ja seurausta esimerkiksi siitä, että kaikista verbeistä ei ole mahdollista muodostaa *-mA-* tai *-in-*loppuisia yhdyssosamuotoja (erilaiset rajoitukset ovat tavallisia kaikessa johtamisessa, ks. esim. Rintala 1978). Aina vaatimukset sanaluokan muuttamisesta eivät ole ehdottomia, vaan ne kumpuavat käytettävyydestä. Esimerkiksi infinitiivi- ja partisiippimuodot sekä jotkin adjektiivit voivat tuntua käytössä hankalilta ja kankeilta, ja todennäköisesti siksi niitä edustaa yhdyssanan alkuosana usein substantiivi (esim. *kaaritamponi* 'kaareva tamponi').

Selvien substantiivialkuosien lisäksi tutkimusaineistossa on 743 sellaista kaksisosaista yhdyssanaa, joiden alkuosa on edusosasta viivalla erotettu sitaattilaina (esim. *auyrveda-lääkintä*, *leprechaun-haltija*; näistä 99:n alkuosa on vieraskielinen sanaliitto, joka on katsottava suomen kielessä yhdeksi leksikaaliseksi osaksi, esim. *tom yum -tahna*, *hands-on-johtaminen*), ja 554 sellaista yhdyssanaa, joiden alkuosa on lyhenne (*ADHD-aikuinen*, *nimby-niuhottaja*). Sitaattilaina- ja lyhennealkuiset sanat on jätetty osa-aineiston 3 ulkopuolelle (osa-aineistot on esitelty luvussa 4.3), koska niiden alkuosan sanaluokan ja morfologisen muodon määrittäminen on tulkinnanvaraista ja niiden mukaan ottaminen voisi vääristää tuloksia. Sitaattilainat voivat lähtökielellä edustaa muutakin sanaluokkaa kuin nomineja (esim. *cheer-tanssi*, *check in -kone*) mutta suomeen lainattuina ne ovat ensisijaisesti asioiden nimityksiä ja siten yhdyssanan alussa nominatiivisijaisia substantiiveja. Myös lyhenteiden on yleensä katsottava edustavan yhdyssanassa ensisijaisesti nominatiivimuotoista substantiivia, vaikka niiden taustalla voi olla mitä moninaisimpia rakenteita (esim. *nimby < not in my backyard*).

Sanaluokaltaan epäspesifien alkuosien ryhmään kuuluvat alkuosat eivät esiinny itsenäisinä eikä niillä siten ole omaa sanaluokkaa. Sanaluokaltaan epäspesifien alkuosien ryhmästä suurimman joukon muodostavat prefiksinkaltaiset alkuosat (*etäopettaa* ja *antimainonta*; prefiksinkaltaisista alkuosista ks. tarkemmin luku 5.3.2.2). Niiden lisäksi tähän ryhmään kuuluvat myös muut selvästi mitään itsenäistä sanaa edustamattomat epäitsenäiset alkuosat (*alkeishiukkanen*, *vihersilta*) sekä selvästi mitään tiettyä sanaluokkaa edustamattomat sanavartalot (esim. *lököhousut*).

Substantiivien ja sanaluokaltaan epäspesifien jälkeen seuraavaksi yleisin uusien yhdyssanojen alkuosan sanaluokka on adjektiivi. Niitä seuraavat useaa sanaluokkaa edustavat alkuosat. Ryhmään kuuluu yhteensä 714 tapausta, joista 538 on substantiivi-adjektiiveja ja 176 muita useaa sanaluokkaa edustavia alkuosia. Samaan tapaan kuin kokonaisissa muodosteissa myös alkuosissa tyypillisimpiä substantiivi-adjektiiveja ovat värinimet sekä ihmisten ja ihmisryhmien nimitykset (*harmaaviini*, *irkkutanssi*). Sitä, kumpi alkuosan sanaluokkapotentiaalissa mahdollisista sanaluokista kussakin yhdyssanassa aktuaalistuu, ei ole analysoitu. Näin ollen tähän ryhmään kuuluvat yhtä lailla esimerkiksi sanat *punahallinto* ja *punamullata*. Muiden useaa sanaluokkaa edustavien alkuosien joukkoon kuuluu myös sanoja, jotka voivat

edustaa useampaa kuin yhtä pääsanaluokkaa (esim. sekä pronomininä että adverbina ja partikkelina toimiva *itse*, ISK § 769). Useaa sanaluokkaa edustavia ovat myös sellaiset sanavartalot, joista on muodostettu tai joista voidaan muodostaa eri sanaluokkiin kuuluvia sanoja (esim. *painitaata*, ks. myös ISK § 250). Poikkeuksia ovat kuitenkin sellaiset sanat, joissa eri sanaluokkia edustavat johdokset ovat leksikaalisuudet eri merkityksiin ja yhdyssanassa on selvästi käytetty merkityksistä jompaa-kumpaa. Esimerkissä 20 yhdyssanan *juonimedia* alkuosa on substantiivisessa merkityksessä, jossa sitä ei voi tulkita liittyväksi juonimiseen vaan ainoastaan juoniin merkityksessä 'kirjallisen tuotteen, elokuvan, tarinan tms. toiminnan kulku, tapahtumakulku' (KS s. v. *juoni*).

- (20) Elokuva ei ole *juonimedia*. Harvoja poikkeuksia lukuun ottamatta juoni on täysin selvä jo elokuvan alkaessa. [TT <- Savon Sanomat, 22.5.2001, s. 27]

Taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvista sanoista uusien yhdyssanojen alkuosana esiintyy sekä adverbijä, adpositioita että partikkeleja (esim. *väkisinmyynti*, *vastavammautunut*, *anteeksikone*). Koska yhdyssanan alkuosina toimivilla sanoilla ei yleensä ole syntaktisia ominaisuuksia ja koska monet taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvat sanat voivat esiintyä sekä adverbina että adpositioina (ISK § 678–679), kaikkia taipumattomia ja vajaasti taipuvia sanoja on tässä luvussa käsitelty yhtenä ryppäänä. Taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvien alkuosien ryhmä on kirjava, eikä tyypillisiä tapauksia ole helppo nostaa esiin. Usein toistuvia alkuosia ovat kuitenkin esimerkiksi *läpi-*, *myötä-* ja *poikki-* (samojen lekseemien toistumisesta ks. luku 6.2).

Verbejä edustaviksi yhdyssanan alkuosiksi olen laskenut verbin *-mA-* ja *-in-*loppuiset yhdysosamuodot (esim. *syömähäiriö*, *kuolinapu*; muodoista ks. tarkemmin luku 5.3.2.3) sekä sellaiset verbivartalot, jotka eivät vastaa suomen kielessä käytössä olevaa nomina (esim. *leijukattila*, *rääkymarkkinointi*; vrt. em. *hyökylurmi*, *painitaata*). Verbialkuisten yhdyssanojen vähäisyys selittyy osittain sillä, että suomen kielessä persoonamuotoisia verbejä ei voi käyttää yhdyssanan alkuosana ja että läheskään kaikista verbeistä ei voi luontevasti muodostaa *-mA-* tai *-in-*loppuisia yhdysosamuotoja. Toinen syy verbialkuosien vähyydelle on se, että *-mA-* ja *-in-*yhdysosamuotoja ei helposti käytetä sellaisista sanoista, joista ne eivät ole vanhaan olleet käytössä. ISK:n (§ 417) mukaan *-mA-*tyyppi tosin on karttuva, kun taas *-in-*tyyppi ei (poikkeuksista ks. kuitenkin luku 5.3.2.3).

Pronominien ryhmässä jopa 71,3 % (72/101 kpl) tapauksista on kvanttoripronominin *moni-* esiintymiä (*monisairas*, *moniuskontoinen*). Merkityksensä kannalta *moni-*alkuosa vastaa pitkälti prefiksinkaltaista *multi-*elementtiä. Koska *moni-*alkuosalla on kuitenkin myös selvästi pronominaalisia merkityksiä ja se vastaa muodoltaan pronomina, sen esiintymät on tässä tutkimuksessa laskettu pronomineiksi. Pronominien kohdalla vaikuttaa siltä, että vain tietyt lekseemit esiintyvät produktiivisesti uusien yhdyssanojen alkuosina – näistä tyypillisimpänä juuri *moni-*. ISK:n jaon mukaisesti pronomineiksi on laskettu myös sellaiset *kuin-*rakenteessa

esiintyvät sanat kuin *sama, eri, muu* ja *toinen* (ISK § 766). Lisäksi mukana on muutamia persoonapronomineja, joiden poikkeuksellisuutta yhdyssanan alkuosana osoittaa se, että ne on usein erotettu jälkiosasta viivalla (*me-tunne, minä-ohjautuva*; ks. myös Eronen 2012: 40).

Numeraalialkuosien luokkaan olen laskenut sekä kirjaimin että numeroin kirjoitetut ordinaali- ja järjestysluvut sekä niiden erilaiset yhdysosamuodot. Numeraaleiksi ei ole laskettu numeraalin substantiivijohdoksia (esim. yhdyssanoissa muodossa *ykkös-* esiintyvä *ykkönen*) tai niiden puhekielisiä variantteja (esim. *kasi-*) (ISK § 791). Kirjaimin kirjoitetuista numeraalialkuosista yleisimmin esiintyvät perusluvut nolasta viiteen (esim. *nollaosinko, yksipyöräinen, kaksikäinen*).

Kaikki aineistossa yhdyssanan alkuosina esiintyvät partisiipit ovat leksikaalistumattomia, ja sikäli myös ne voidaan katsoa verbiä edustaviksi (esim. *nuokkuvakukkainen, ummistettusilmäinen*). Aineiston partisiippialkuiset yhdyssanat ovat yhtä lukuun ottamatta *-inen*-loppuisia adjektiiveja, jotka ovat syntyneet sanaliitoista tiivistymällä tilanteessa, jossa edusosasta on muodostettu *-inen*-johdos. Myös partisiipin ja substantiivin yhdistäminen on mahdollista (*naituahneus* 'avioliiton kautta sukuun tullut ahne ihminen t. tällaisen ihmisen luonteenpiirre [esim. perinnönjakotilanteessa]'). Substantiiviloppuisissa sanoissa käytetään kuitenkin partisiipin merkityksessä yleisimmin sanavartaloa tai sen kanssa identtistä substantiivia (esim. *riippupeppu, liukupysähdys*).

Alkuosan sanaluokkaa aiemmin käsitellyt Jussi Niemi (2009: 251) esittelee CD-PS:n yhdyssanojen alkuosan sanaluokkajakaumaa koko muodosteen sanaluokan mukaan jaoteltuna (yhdyssubstantiivit, -adjektiivit ja -verbit). Uuden yhdyssanaston ja CD-PS:n yhdyssubstantiivien alkuosan sanaluokkajakaumissa ei ole tilastollisesti merkitseviä eroja (Fisherin tarkan testin $n = 2\ 000$, $p = 0,4$). Sen sijaan sekä yhdysadjektiivien että yhdysverbien luokissa uuden yhdyssanaston alkuosien sanaluokkajakauma poikkeaa Fisherin tarkan testin mukaan tilastollisesti merkitsevästi CD-PS:n jakaumasta (adjektiivit: $n = 1\ 195^{69}$, $p < 0,001^{***}$, verbit: $n = 611$, $p < 0,001^{***}$).

Sovitettujen standardoitujen jäännösten avulla tehdyn jatkotarkastelun mukaan yhdysadjektiivien ryhmien välillä oleva tilastollisesti merkitsevä ero on seurausta siitä, että uusien yhdysadjektiivien joukossa on CD-PS:aa enemmän substantiivialkuisia ja sanaluokaltaan epäspesifin alkuosan sisältäviä tapauksia (sovitettut standardoidut jäännökset: substantiivialkuiset = 4,3, sanaluokaltaan epäspesifillä alkuosalla alkavat = 3,5) mutta vähemmän adjektiivi- ja numeraalialkuisia tapauksia (sovitettut standardoidut jäännökset: adjektiivialkuiset = -7,2, numeraalialkuiset = -3,6). Myös yhdysverbien joukossa on uudessa yhdyssanastossa tilastollisesti

69 Jussi Niemen (2009: 251) analysoimassa aineistossa ovat mukana kaikki CD-PS:n yhdysverbit, mutta yhdyssubstantiiveista ja -adjektiiveista mukana on vain noin 5 % kaikista tapauksista sisältävä satunnaisotos. Vertailua varten olen rajannut omaa tutkimusaineistoani Niemen (2009: 250–251) käyttämällä kriteereillä. Käytetyssä aineistossa ovat mukana kaikki Niemen kriteereihin sopivat adjektiivit ja verbit (910 adjektiivia ja 294 verbiä omasta aineistostani sekä 285 adjektiivia ja 317 verbiä Niemen aineistosta).

merkitsevästi CD-PS:aa enemmän substantiivialkuisia tapauksia (sovitettut standardoidut jäännös = 3,1) ja vähemmän adjektiivi- ja numeraalialkuisia tapauksia (sovitettut standardoidut jäännökset: adjektiivialkuiset = -2,5, numeraalialkuiset = -3,6). Erot muita sanaluokkia edustavien alkuosien määrissä eivät ole tilastollisesti merkitseviä.

Uuden yhdyssanaston ja CD-PS:n yhdyssanojen (Jussi Niemi 2009: 251) välillä olevat tilastolliset erot osoittavat selvästi, että suomen kielen uusien yhdysadjektiivien ja -verbien sekä sanakirjoihin otetun vakiintuneen sanaston välillä on rakenteellisia eroja. Substantiivialkuosien suurempi määrä uudessa yhdyssanastossa viittaa yhdyssanojen rakenteen yhdenmukaistumiseen. Kaikkien sanaluokkien yhdyssanoissa on vanhastaankin ollut eniten substantiivialkuisia tapauksia, mikä on omiaan lisäämään substantiivialkujen käyttöä uudessa sanastossa entisestään (produktiivisuuden muutoksen syistä ks. tarkemmin luku 3.2.3). Substantiivialkujen ja sanaluokaltaan epäspesifien alkuosien lisääntyessä muihin sanaluokkiin kuuluvien alkuosien osuudet vastaavasti pienenevät. Sekä yhdysadjektiivien että yhdysverbien ryhmissä voimakkaimmin vähenevät adjektiivi- ja numeraalialut. Sanaluokaltaan epäspesifin alkuosan sisältävien yhdyssanojen lisääntymisen suurin syy on todennäköisesti prefiksinkaltaisten yhdysosien kasvava suosio (ks. tarkemmin luku 5.3.2.2; ks. myös Häkkinen–Björkqvist 1991: 52). Niiden lisääntymistä ei selitä niiden aiempi yleisyys suomen kielessä, mutta vieraiden kielten tarjoamalla malleilla voi olla vaikutusta suosioon. Lisäksi prefiksinkaltaiset alkuosat pystyvät luontevasti täyttämään sellaisia ilmaisutarpeita, jotka voisivat muuten vaatia kankeita tai pitkiä rakenteita. (Ks. myös O. Nuutinen 1989: 118.)

5.2.5 Yhteenveto

Suurin osa suomen kielen uusista yhdyssanoista on substantiiveja. Niiden osuus on uudessa yhdyssanastossa suurempi kuin sanakirjoihin vakiintuneissa yhdysanoissa. Mahdollinen selitys tälle on esimerkiksi se, että erikoiskielisen viestinnän suuntaan muuttuvassa yleiskielessä tarvitaan paljon alakäsitteiden nimityksiä, joiksi yhdyssubstantiivit luontevasti sopivat. Myös yhdysverbien osuus on uudessa yhdyssanastossa suurempi kuin sanakirja-aineistoissa. Tämä voi kuitenkin olla seurausta siitä, että yhdysverbejä on tietoisesti jätetty sanakirjojen ulkopuolelle. Joka tapauksessa yhdysverbien runsaus aineistossa osoittaa, että yhdysverbien kategoria on suomessa produktiivinen ja että yhdysverbejä pidetään yleiskielessä hyödyllisenä ilmaisukeinona. Yhdysadjektiivien osuus puolestaan on uudessa yhdyssanastossa pienempi kuin sanakirja-aineistoissa. Yhdysadjektiivien vähyys ei näytä johtuvan siitä, että ns. muotiloppuja olisi alettu välttää.

Yhdyssanan kontekstisanaluokka suomen kielessä aina sama, jossa sen jälkiosa kyseisessä kontekstissa esiintyy. Yhdyssanan potentiaalinen sanaluokka on kuitenkin enemmän seurausta yhdyssanan osien yhdessä luomasta merkityksestä kuin pelkän jälkiosan potentiaalisesta sanaluokasta. Yhdyssanan alkuosa voi rajata jälkiosan potentiaalisista sanaluokista mahdolliseksi vain jonkin. Jälkiosan ja koko muodosteen sanaluokkapotentiaali voi olla erilainen varsinkin sellaisissa sanoissa,

joissa semanttinen pääjäsen on poikkeuksellisesti yhdyssanan alkuosa. Eksosentristen yhdyssanojen sanaluokkapotentiaali voi olla erilainen kuin niiden kummankaan osan.

Suomen kielen uusista yhdyssanoista valtaosa on substantiivialkuisia. Niiden jälkeen yleisyysjärjestyksessä seuraavia ovat sanaluokaltaan epäspesifit alkuosat, adjektiivit sekä useaa sanaluokkaa edustavat alkuosat. Muita sanaluokkia edustavia alkuosia on tutkimusaineistossa alle 1 % kutakin. Yhdysubstantiiveista lähes 90 % on substantiivialkuisia sekä sanakirja-aineistossa että uudessa yhdyssanas-tossa. Sanakirja-aineistoon verrattuna substantiivialkuisuus on yleisempää myös yhdysadjektiiveissa ja yhdysverbeissä. Yksi substantiivien ylivallan syy on se, että ne edustavat yhdyssanojen alkuosissa usein myös verbaalisia ja adjektiivisia merkityksiä. Yhdysadjektiivien ryhmässä sanaluokaltaan epäspesifit alkuosat ovat sanakirjoihin verrattuna yleisempiä. Niistä suurin osa on prefiksinkaltaisia elementtejä.

5.3 Alkuosan muoto

Tässä luvussa käsittelen suomen kielen uusien yhdyssanojen alkuosan morfologiasta muotoa kahdesta eri näkökulmasta. Kahdessa ensimmäisessä alaluvussa tarkastelen tiettyyn muotoon kiteytyneitä yhdyssanan alkuosia ja niiden muotojakaumaa. Käytän kiteytyneiden alkuosien morfologisen muodon analyysissa kahta luokittelutapaa. Ensin esittelen karkean jaottelun, jonka tarkoitus on antaa yleiskuva aineistosta sekä mahdollistaa vertailu aiempiin tutkimuksiin (5.3.1). Toisessa vaiheessa analysoin hienojakoisemman luokittelun avulla karkeassa luokittelussa esiteltyjen ryhmien sisäistä variaatiota tarkemmin (5.3.2). Kolmannessa alaluvussa (5.3.3) käsittelen sellaisia yhdyssanan alkuosia, jotka kongruoivat yhdyssanan jälkiosan kanssa ja joiden muoto näin ollen vaihtelee. Jälkiosan kanssa kongruoivien alkuosien harvinaisuuden vuoksi kolmas alaluku keskittyy kvalitatiiviseen analyysiin.

5.3.1 Kiteytyneen alkuosan muoto: karkea luokittelu

Yhdyssanan alkuosan muoto on olennainen tarkastelukohde informaation tiivistämisen ja sananmuodostuksessa käytettävän rakenteen valinnan kannalta. Nominatiivi jättää yhdyssanan tulkintamahdollisuudet hyvin avoimiksi, kun taas muilla muodoilla merkitystä voidaan rajata tarkemmin (ks. tarkemmin luku 2.3.2). Nominatiivialku vaikuttaa valtaavan alaa, joten keskeinen kysymys on, kuinka paljon suomen uudessa yhdyssanas-tossa ylipäänsä käytetään muita alkuosan muotoja.

Tässä luvussa on käytetty osa-aineistoa 3, joka sisältää vain kaksiosaisia yhdys-sanoja ($n = 22\ 144$, osa-aineistot on esitelty tarkemmin luvussa 4.3). Tiedot tutkimusaineiston yhdyssanojen alkuosan karkeasta muotojakaumasta on koottu taulukkoon 6.

TAULUKKO 6. Tutkimusaineiston yhdyssanojen alkuosan karkea morfologinen muotojakauma ($n = 22\ 144$)

Alkuosan muoto	Esimerkki	Osuus	Määrä
Nominatiivi	<i>hattufriikki, kansalaisuustesti</i>	79,5 %	17 599
Epäitsenäinen elementti	<i>antimainonta, etäpankki</i>	9,3 %	2 052
Yhdysosamuoto	<i>ihmisbanaani, kellumahiiri</i>	6,1 %	1 359
Genetiivi	<i>hiekanveistäjä, piispankaupunki</i>	3,9 %	865
Taipumaton tai vaillinaisesti taipuva sana	<i>kanssakädellinen, päättökastelu</i>	1,1 %	241
Muu sijamuoto	<i>maallemuutto, työssäoppiminen</i>	0,1 %	28
Kaikki muodot		~ 100 %	22 144

Kuten taulukosta 6 nähdään, uudessa yhdyssanastossa ylivoimaisesti yleisin alkuosan muoto on nominatiivi (79,5 % aineistosta). Sen jälkeen yleisimpiä ovat mitään tiettyä itsenäistä lekseemiä edustamattomat epäitsenäiset leksikaaliset elementit (9,3 %, esim. prefiksinkaltaiset alkuosat) ja ainoastaan sidonnaisina esiintyvät itsenäisten sanojen yhdysosamuodot (6,1 %, esim. konsonanttivartaloa vastaavat *-nen*-loppuisten sanojen *-s*-loppuiset yhdysosamuodot). Neljänneksi yleisin alkuosan muoto on genetiivi (yksikkö ja monikko yht. 3,9 %). Taipumattomia ja vaillinaisesti taipuvia sanoja on käytetty alkuosana 1,1 %:ssa aineiston sanoja. Täyden taivutusparadigman omaavista sanoista muita sijamuotoja kuin nominatiivia ja genetiiviä on käytetty alkuosana vain 0,1 %:ssa.

Nominatiivien ylivaltaa uudessa yhdyssanastossa selittää parhaiten se, että se on jo vanhastaan ollut käytetyin alkuosan muoto suomen kielen yhdyssanoissa. Nominatiivialkujen yleisyydestä on huomautettu jo 1800-luvun kieliopeissa, vaikka monet mainitsevat määriteosan olevan hyvin usein myös genetiivissä (esim. Genetz 1881: 21–22; Hämäläinen 1883: 122–124; Setälä 1898: 126–127). Nominatiivimuotoisen alkuosan suosio johtuu todennäköisesti sen rajattomista käyttömahdollisuuksista (ks. myös luvut 2.3.1 ja 2.3.2). Nominatiivialkujen yleisyys osoittaa, että konteksti ja maailmantieto riittävät takaamaan viestin välittymisen, vaikka rakenne itsessään ei anna vihjeitä osien välisistä merkityssuhteista.

Yhdyssanan alkuosana esiintyvistä epäitsenäisistä elementeistä valtaosa on prefiksinkaltaisia elementtejä (ks. tarkemmin 5.3.2.2). Viimeistään tämä tutkimus osoittaa, että prefiksinkaltaiset elementit on syytä huomioida paremmin suomen yhdyssanojen muodostusta käsittelevissä tutkimuksissa. Myös erilaiset yhdysosamuodot ovat yhdyssanan alkuosina yleisempiä kuin vanhastaan tärkeinä pidetyt genetiivit. Yhdysosamuodoista tosin suurin osa on *-nen*-loppuisista sanoista mekaanisesti muodostettuja *-s*-loppuisia muotoja, jotka vastaavat merkitystehtäviltään nominatiivialkuosia (ks. tarkemmin 5.3.2.3).

Nykykielen yhdyssanojen alkuosien muotojakaumaa on tutkittu sanakirja-aineistosta myös kvantitatiivisesti (Saukkonen 1973: 338; Jussi Niemi 2009: 252–253), mutta muotojakauman epäjohdonmukaisten ja implisiittiseksi jäävien kuvausten vuoksi vertailu aiempien tutkimusten tuloksiin on vaikeaa kaikkien yhdyssanojen

ryhmässä. Saukkosen (mp.) tutkimus tarjoaa kuitenkin hyvän kuvan yhdysadjektiivien alkuosien muotojakaumasta NS:ssa.

Uusien yhdysadjektiivien ja NS:n yhdysadjektiivien (Saukkonen 1973: 338) alkuosan muotojakaumat eroavat tilastollisesti merkitsevästi toisistaan (Pearsonin khiin neliö -riippumattomuustestin $\chi^2 = 50,8$, $n = 1\,458^{70}$, $df = 2$, $p < 0,001^{***}$). Uudessa yhdyssanastossa on enemmän nominatiivialkuisia yhdysadjektiiveja ja vähemmän genetiivillä ja muilla sijamuodoilla alkavia yhdysadjektiiveja. Uudessa yhdyssanastossa nominatiivialku on 88,7 %:ssa yhdysadjektiiveista, kun NS:ssa vastaava osuus on Saukkosen (mp.) mukaan noin 75 %. Genetiiviä edustava alkuosa on uudessa yhdyssanastossa vain 11 %:ssa adjektiiveista, kun NS:ssa vastaava osuus on noin 22 % (mp.). Muita sijamuotoja edustavia alkuosia on uudessa yhdyssanastossa 0,3 %:ssa ja NS:ssa noin 3 %:ssa (mp.) yhdysadjektiiveista. Erot ovat sovitettujen standardoitujen jäännösten perusteella tilastollisesti merkitseviä kaikissa alkuosan muotoryhmissä (sovitetut standardoidut jäännökset: nominatiiveissa = 6,6, genetiiveissä = -5,2, muita sijamuotoja edustavissa alkuosissa -4,6).

Vaikka osa erosta voi olla seurausta erilaisista analyysimenetelmistä, näyttää siltä, että uusien yhdysadjektiivien alkuosina todella esiintyy huomattavasti sanakirja-aineistoa vähemmän muita sijamuotoja kuin nominatiivia. Oletusta tukee myös vertailu Jussi Niemen (2009: 252–253) tutkimukseen, jonka mukaan CD-PS:n yhdysadjektiivien alkuosista yhteensä yli 21 % on muussa sijamuodossa kuin nominatiivissa (Niemen käyttämät kategoriat ovat nominatiivi, genetiivi, muu sijamuoto ja taivutettu johdos). Sanakirjoihin vertailun perusteella vaikuttaa siis siltä, että nominatiivialku valtaa alaa muilta alkuosan muodoilta ainakin yhdysadjektiivien ryhmässä. Tarkastelen yhdyssanan alkuosan muodon ja koko muodosteen sanaluokan yhteyttä omassa aineistossani tarkemmin luvussa 7.2.

5.3.2 Kiteytyneen alkuosan muoto: hienojakoinen luokittelu

Siirryn seuraavaksi tarkastelemaan uuden yhdyssanaston alkuosien muotojakaumaa hienojakoisen luokittelun avulla. Luokittelun pohjana on edellä esitelty karkea jaottelu (5.3.1), jonka ryhmiä puran seuraavassa yksitellen pienempiin osiin. Suomen kielen yhdyssanojen alkuosan muotoa ei ole tutkittu aiemmin vastaavalla tarkkuudella. Suurin osa tässä luvussa käytetyistä alkuosan muotokategorioista on kuitenkin mainittu aiemmassakin tutkimuksessa (esim. Vesikansa 1989: 221–229; ISK § 403, 406, 409–419). Mukana on kuitenkin myös yksi kokonaan uusi kategoria (*-inen*-loppuisen vierassanan *-i*-loppuinen yhdysosamuoto, esim. *virtuaalihyödyke*, ks. luku 5.3.2.3). Luvussa on käytetty samaa osa-aineistoa kuin luvussa 5.3.1 (kaksiosaisia yhdyssanoja sisältävä osa-aineisto 3, $n = 22\,144$, osa-aineistot on esitelty tarkemmin luvussa 4.3).

70 Saukkosen analyysin päätarkoitus on yhdyssanojen semanttinen ryhmittely, eikä hän käytä lainkaan taipumattomien ja vajaasti taipuvien alkuosien eikä epäitsenäisten alkuosien kategorioita. Nämä kategoriat on siksi jätetty kokonaan tämän vertailun ulkopuolelle. Testauksessa on käytetty kaikkia oman tutkimusaineistoni nominatiivissa, genetiivissä tai muussa sijassa olevalla alkuosalla alkavia yhdysadjektiiveja (739 kpl) sekä Saukkosen aineistosta poimittuja yhtä suuria satunnaisotoksia ($n = 2 \times 739 = 1\,458$).

5.3.2.1 Nominatiivit

Tutkimusaineiston yhdyssanoista 79,5 % (17 599 kpl) on nominatiivialkuisia. Tiedot nominatiivikategorian sisäisestä jakaumasta on koottu taulukkoon 7.

TAULUKKO 7. Tutkimusaineiston yhdyssanojen nominatiivialkuosien muotojakauma

Alkuosan muoto	Esimerkki	Osuus nominatiivialkuista	Osuus kaikista alkuosista	Lukumäärä
Yksikkö	<i>hotellijuusto, ratkaisukristitty</i>	98,3 %	78,2 %	17 305
Nominatiivi/verbin vokaalivartalo	<i>hyötykävellä, pilkkivaate</i>	1,0 %	0,8 %	180
Monikkosana yksikössä	<i>aivotuhlaus, talkooverkko</i>	0,6 %	0,5 %	106
Partisiippi	<i>kapenevakärkinen ummistettusilmäinen</i>	0,04 %	0,03 %	7
Monikko	<i>aikojat-lista</i>	0,006 %	0,005 %	1
Yhteensä		~ 100 %	~ 79,5 %	17 599

Kuten taulukosta 7 nähdään, nominatiivimuotoisten alkuosien ylivoimaisesti yleisin tyyppi on tavallisten perussanojen ja johdosten yksikkömuoto (98,3 % kaikista nominatiivialkuista). Paljon niitä harvinaisempia ovat alkuosat, jotka voidaan katsoa joko nominatiivimuotoisen nominin tai verbin vartalon edustumiksi (*hehkusytytin, keuhkala*; 1 % kaikista nominatiivialkuista). Jos alkuosa esiintyy itsenäisenä nominina, se on luokiteltu analyysissä tähän kategoriaan. Jos taas alkuosaa ei käytetä itsenäisenä nominina, se on laskettu verbin vartaloksi (esim. *riippupeppu*, käsittelen tällaisia epäitsenäisiä yhdyssosamuotoja tarkemmin luvussa 5.3.2.3).

Morfologiselta muodoltaan nominatiiveja ovat myös monikko- eli *plurale tantum* -sanojen yksikkömuodot (0,6 % kaikista nominatiivialkuisista). Nämä sanat esiintyvät itsenäisinä vain monikossa (esim. *aivot, häät*), mutta yhdyssanan alkuosana toimiessaan ne on perinteisesti tulkittu nominatiiveiksi (*aivoproteesi, kasvotunnistus*) (Setälä 1898: 126–127; ISK § 558). Koska nämä muodot eivät esiinny itsenäisinä, ne voitaisiin periaatteessa tulkita myös epäitsenäisiksi yhdyssosamuodoiksi (ks. luku 5.3.2.3).

Harvinaisimpia yhdyssanan alkuosan nominatiivimuotoisia ryhmiä ovat partiisiipit (*nuokkuvakukkainen, naituahneus* 'avioliiton kautta sukuun tullut ahne ihminen t. tällaisen ihmisen luonteenpiirre [esim. perinnönjakotilanteessa]') sekä monikon nominatiivi (molempia alle 0,1 % kaikista nominatiivialkuisista). Monikon nominatiivi voi ISK:n (2004: § 418) mukaan esiintyä yhdyssanan alkuosana vain jälkiosan ollessa partiisiippi tai sen johdos (esim. *veretseisauttava*). Omassa aineistossani monikon nominatiivi esiintyy vain yhdessä muodosteessa, jonka jälkiosa ei kuitenkaan ole partiisiippi tai sen johdos (esim. 21).

- (21) Tamperelainen ohjelmistoyhtiö Analyste ilmoitti eilen poistuvansa Helsingin pörssin aikojien listalta, joka on puolivälin pysäkki matkalla

pörssiin. – – Aktia on päättänyt pysyä *aikoajat-listalla* eli tavallaan listautumisen puolivälissä. [TT <- HS 31.5.2002, s. D2]

Poikkeuksellinen rakenne selittyy alkuosan nimimäisyydellä. Rakennetta tuskin voi pitää produktiivisena yhden tapauksen perusteella, mutta täysin mahdoton se ei ole. Vastaavaa rakennetta noudattavaa yhdyssanaa *muut-kategoria* on käytetty myös tässä tutkimustekstissä (ks. luvut 5.3.2.2 ja 5.3.2.3).

5.3.2.2 Epäitsenäiset leksikaaliset elementit

Tutkimusaineistoni yhdyssanoista 9,3 % (2 052 kpl) sisältää mitään itsenäistä lekseemiä edustamattoman alkuosan. Tiedot epäitsenäisten elementtien kategorian sisäisestä jakaumasta on koottu taulukkoon 8. Toisin kuin suurin osa tämän tutkimuksen luokitteluista, epäitsenäisten elementtien jaottelu ei perustu kaikilta osin pelkkään osien muotoon, sillä prefiksinkaltaisten elementtien erottamisessa muista on ollut välttämätöntä kiinnittää huomiota myös elementin merkitykseen, käyttöön ja alkuperään.

TAULUKKO 8. Tutkimusaineistossa yhdyssanan alkuosana toimivien epäitsenäisten elementtien muotojakauma ($n = 2\ 052$)

Alkuosan muoto	Esimerkki	Osuus epäitsenäisistä	Osuus kaikista alkuosista	Lukumäärä
Kotoinen prefiksinkaltainen elementti	<i>epäsoitin, peruskompastelu</i>	48,6 %	4,5 %	998
Vierasperäinen prefiksinkaltainen elementti	<i>antimainonta, geokätkö</i>	44,9 %	4,2 %	922
Muu (esim. uniikkivartalo tai ainoastaan yhdyssanan alkuosana esiintyvä johdos)	<i>sipiselvä, kaksoiskriisi, kukkaisumous</i>	6,4 %	0,6 %	132
Yhteensä		~ 100 %	~ 9,3 %	2 052

Epäitsenäisistä elementeistä yleisimpiä ovat kotoiset ja vierasperäiset prefiksinkaltaiset alkuosat (kotoiset 48,6 % ja vierasperäiset 44,9 % kaikista epäitsenäisistä elementeistä). Prefiksinkaltaisten elementtien lisäksi epäitsenäisten elementtien joukossa on esimerkiksi ainoastaan yhdyssanan alkuosana esiintyviä johdoksia ja uniikkivartaloita (muut epäitsenäiset elementit yhteensä 6,4 % kaikista epäitsenäisistä elementeistä).

Kielissä joissa on prefiksijärjestelmä, sananmuodostuksen teoriassa erotetaan yleensä prefiksit (esim. *re-*, *anti-*), jotka muodostavat sanaan liittyessään johdoksen, ja yhdyssanan alkuosamuodot⁷¹ (*initial combining form*), jotka sanaan liittyessään muodostavat yhdyssanan (esim. *agri-/agro-*, *geo-*; kriteereistä tarkemmin esim.

71 Käsite on huomattava erottaa fennistiikassa käytetystä yhdysosamuodon käsitteestä. Yhdysosamuodolla tarkoitetaan itsenäisestä sanasta yhdyssanan alussa käytettävää erityistä epäitsenäistä muotoa, kun taas yhdyssanan alkuosamuodot (*initial combining forms*) eivät välttämättä ole minkään (niitä hyödyntävän kielen leksikkoon kuuluvan) itsenäisen sanan muotoja.

Prčić 2005; L. Kolehmainen 2006: 3). Tällaiset usein alkujaan esimerkiksi kreikkalais-, latinalais- ja englantilaisperäisistä sanoista tyypistämällä syntyneet alkuosat ovat yleisiä yhdyssanan alkuosia monissa kielissä (esim. *Europa* > *euro-*, *digital* > *digi-*) (Booij 2007: 90). Näiden kahden ryhmän johdonmukainen erottaminen on kuitenkin vaikeaa samaan tapaan kuin johdinten ja perussanojen erottaminen suomessa (esim. *-läntä*, ks. luku 2.1.2; englannin osalta esim. Prčić 2005). Koska suomen kielessä ei ole varsinaista prefiksisysteemiä, varsinaisten prefiksien ja prefiksinkaltaisten yhdyssanan alkuosamuotojen erottaminen ei ole suomen sananmuodostuksen tutkimuksessa tarpeen (sekä muista kielistä lainautuneiden prefiksien että yhdyssanan alkuosamuotojen avulla muodostetaan suomessa yhdyssanoja, ei johdoksia⁷²; ks. esim. Häkkinen–Björkqvist 1991: 51–52). Suomen sananmuodostuksessa lähtökielessä eri tavalla käyttäytyvien elementtien eroksi jääkin lähinnä merkityksen tarkkuus: lähtökielen prefiksit ovat merkitykseltään yleisluonteisempia kuin yhdyssanan alkuosamuodot. Sananmuodostuksen kannalta näiden kategorioiden välillä ei ole eroa, sillä molemmat ovat epäitsenäisiä, vierasperäisiä yhdyssanan alkukomponenttina toimivia elementtejä. Jatkossa kutsun vieraissa kielissä sekä prefiksinä että yhdyssanan alkuosamuotona pidettyjä elementtejä yhteisellä nimellä *prefiksinkaltaiset elementit*. Nimitys kattaa myös kotoperäiset prefiksimäiset elementit (esim. *epä-*, *lähi-*).

Esimerkiksi Olli Nuutisen (1989: 110–119) mukaan malli etuliitteiden käyttöön on saatu suomeen käännöslainauksen kautta muista Euroopan kielistä, ja myös osaa kotoperäisistä suomen kielen yhdyssanojen epäitsenäisistä alkuosista on toisinaan nimitetty prefiksinkaltaisiksi (esim. ISK § 172). Esimerkiksi ISK:ssa (§ 172) prefiksiksi määritellään sanavartalon alkuun liittyvät sidonnaiset morfeemit, jotka rinnastuvat tehtäviltään johtimiin ja yhdysosiin. Varsinaisiksi omaperäisiksi prefiksiksi ISK:ssa mainitaan kieltoprefiksit *epä-* ja *ei-*. Sen sijaan sellaisia elementtejä kuin *lähi*, *nyky-*, *täsmä-* jne. nimitetään ISK:ssa nominaalisiksi vartaloiksi ja prefiksimaaisiksi aineksiksi (ks. myös ISK § 305).

Tässä tutkimuksessa tehdyt rajaukset poikkeavat jonkin verran prefiksinkaltaisista elementeistä aiemmin kirjoitetusta. Käytännössä eroja on kuitenkin vain yksittäisten elementtien luokittelussa. Oman luokitteluni taustalla vaikuttavat sekä alkuosan muoto että merkitys. Jos sana esiintyy myös itsenäisenä eikä sen merkitys itsenäisenä sanana poikkea olennaisesti sen merkityksestä yhdyssanan alkuosana, olen tulkinnut sanan itsenäiseksi lekseemiksi. Esimerkiksi ISK:n (§ 172) epäitsenäiseksi alkuosaksi mainitsema *liika* esiintyy yhdyssanoissa yleensä vastaavassa merkityksessä kuin itsenäisenä adjektiivinakin (vrt. *liika tankkaus johti lämmönsäätelyn häiriintymiseen ~ välttä siis liikatankkausta*) ja on siksi luokiteltu tässä tutkimuksessa nominatiivimuotoisten alkuosien ryhmään. Jos alkuosa on selvästi merkitykseltään sanan itsenäisestä käytöstä irtaantunut, olen luokitellut sen prefiksimaaiseksi ainekseksi. Esimerkiksi sellaiset alkuosat kuin *pää-*, *sika-*, *umpi-* olen tulkinnut joko

72 Prefiksinkaltaiset elementit eivät mukaudu vokaalisointuun. Suomen kirjoitussysteemissä prefiksinkaltaiset elementit on huomioitu yhdysosina (yhdyssanoissa samojen vokaalien välissä käytettävä yhdysviiva, esim. *yli-ihminen*).

nominatiivimuotoisiksi nomineiksi tai epäitsenäisiksi prefiksinkaltaisiksi elementeiksi niiden merkityksen mukaan (vrt. nominatiivialkuiset *päähieronta* 'pään hieronta', *sikarutto* 'sikojen rutto', *umpikypärä* 'umpinainen kypärä' ja prefiksinkaltaisella elementillä alkavat *pääkoti* 'kahdesta kodista tärkeämpi', *sikacool* 'todella cool', *umpirakastunut* 'täysin rakastunut'). Tässä tutkimuksessa kotoisten prefiksinkaltaisten alkuosien ryhmään on laskettu myös kieltoverbi *ei-*, vaikka se on muodoltaan identtinen itsenäisen verbimuodon kanssa. Yhdyssanoissa sen merkitys on kuitenkin prefiksinkaltainen (vrt. *epä-*; myös ISK luokittelee *ei-*elementin prefiksiksi, § 172; *ei-*alkuosasta esim. Eronen 2010).

Prefiksinkaltaisia elementtejä käytetään uudessa yhdyssanastossa ahkerasti. Kotoisia (48,6 % kaikista epäitsenäisistä elementeistä) ja vierasperäisiä (44,9 % kaikista epäitsenäisistä elementeistä) prefiksinkaltaisia elementtejä käytetään osapuilleen yhtä paljon. Prefiksinkaltaisia elementtejä ei ole käsitelty suomen kielen yhdyssananmuodostusta koskevissa tutkimuksissa paljoa, eikä niitä ole huomioitu esimerkiksi tämän tutkimuksen vertailuaineistona käytetyissä sanakirja-aineistoa käsittelevissä tutkimuksissa (Saukkonen 1973; Jussi Niemi 2009). Jo aiemmin Häkkinen ja Björkqvist (1991: 52) ovat kuitenkin huomauttaneet prefiksinkaltaisten yhdysosien käytön muodostavan yhdyssananmuodostuksessa selvän, sekä produktiivisuutensa että funktionsa puolesta yhä tärkeämmäksi muuttuvan alasynteesin. Viimeistään tämä tutkimus osoittaa, että prefiksinkaltaiset elementit ovat vallanneet paikkansa suomen kielen sananmuodostuksessa. Muihin suomen kielen yhdyssanojen alkuosiin verrattuna prefiksinkaltaisten elementtien käytössä on huomionarvoista samojen elementtien tiheä käyttö (ks. tarkemmin luku 6.2).

Muut-kategoriaan kuuluvat alkuosat (6,4 % kaikista epäitsenäisistä elementeistä) ovat sekalainen joukko itsenäistä sanaa edustamattomia elementtejä, jotka eivät kuitenkaan täytä prefiksinkaltaisen elementin kriteerejä: ne eivät esimerkiksi voi toimia johdosten kantoina kuten kotoiset prefiksinkaltaiset elementit tai ne eivät ole vakiintuneet tiheään käyttöön (ks. myös Prčić 2005: 324–327), vaan esiintyvät vain yhden tai korkeintaan muutaman jälkiosan kanssa (esim. ns. uniikkivartalot, ks. ISK § 406). Merkitykseltään muut-kategoriaan kuuluvat alkuosat ovat yleensä tarkempia kuin melko abstraktit prefiksinkaltaiset elementit (prefiksinkaltaisten elementtien merkityksen tarkkuudesta esim. Pilke 1999). Selkeimmin joukosta erottuva ryhmä ovat sellaiset *-nen-*loppuisten sanojen *-s-*loppuisia yhdysosanmuotoja muistuttavat johdokset, jotka esiintyvät ainoastaan yhdyssanan alkuosana (esim. *kaksoismörkö*, *tieteisnörtti*). Tällaisia alkuosia on aineistossa yhteensä 54 kpl (39 tapauksista on *kaksois-*elementin esiintymiä). Epäitsenäisten johdosten lisäksi muut-kategoriaan kuuluvia tapauksia ovat esimerkiksi ns. uniikkivartalot (esim. *maalissää*, *nattapala* 'yöpala') sekä intensifioivat allitteraation kautta muodostetut alkuosat (esim. *paljopaljas*, *sipiselmä*).

5.3.2.3 Yhdysosamuodot

Tutkimusaineistoni yhdyssanoista 6,1 % (1 359 kpl) sisältää alkuosan, joka on itsenäisen sanan epäitsenäinen yhdysosamuoto. Tiedot yhdysosamuotojen sisäisestä jakaumasta on koottu taulukkoon 9. Jaottelu perustuu ensisijaisesti alkuosan muotoon. Merkityksen perusteella muista on erotettu värinimitysten ja numeraalien yhdysosamuodot, joiden muodostus ei ole säännönmukaista.

TAULUKKO 9. Tutkimusaineistossa yhdyssanan alkuosana toimivien yhdysosamuotojen muotojakauma ($n = 1\,359$)

Alkuosan muoto	Esimerkki	Osuus yhdysosamuodoista	Osuus kaikista alkuosista	Lukumäärä
- <i>nen</i> -loppuisen sanan - <i>s</i> -loppuinen konsonanttivartalo	<i>ihmistorni, yleisperuna</i>	57,4 %	3,5 %	780
- <i>inen</i> -loppuista vierassanaa vastaava - <i>i</i> -loppuinen yhdysosamuoto	<i>digitaalipersoona, totaalitiltti</i>	12,1 %	0,7 %	165
Muu kuin - <i>nen</i> -loppuisen sanan konsonanttivartalo	<i>pienlaina, täysautomaattinen</i>	11,1 %	0,7 %	151
Värinimityksen tai numeraalin yhdysosamuoto	<i>kolmilaatikkoinen, viherkaulus</i>	8,0 %	0,5 %	109
Verbin - <i>mA</i> -loppuinen muoto	<i>ampumapaini, syntymälihava</i>	3,1 %	0,2 %	42
Vokaalivartalo (verbin vartalo tai substantiivin monikkovartalo)	<i>kiilugeeni, riippupeppu, sikopaja</i>	2,5 %	0,2 %	34
Verbin - <i>in</i> -loppuinen muoto	<i>asuintorni, kuolinapu</i>	2,2 %	0,1 %	30
<i>s</i> -linkityselementin sisältävä muoto	<i>pappispula, rouvasporno</i>	0,9 %	0,05 %	12
Muu yhdysosamuoto	<i>aluvanne, normiporukka, ottipaikka, pressokone</i>	2,6 %	0,2 %	36
Yhteensä		~ 100 %	~ 6,1 %	1 359

Yhdysosamuodoista selvästi yleisimpiä ovat -*nen*-loppuisen sanojen -*s*-loppuiset konsonanttivartalot (57,4 % kaikista yhdysosamuodoista). Seuraavina tulevat -*inen*-loppuisia vierassanoja vastaavat -*i*-loppuiset yhdysosamuodot (12,1 %), muut konsonanttivartalot (11,1 %), värinimityksen tai numeraalin muodostukseltaan vaihtelevat yhdysosamuodot (8 %), verbin MA-infinitiivin vartaloa vastaavat -*mA*-loppuiset yhdysosamuodot (3,1 %), vokaalivartalot (2,5 %) ja verbin -*in*-loppuiset yhdysosamuodot (2,2 %). *s*-linkityselementin sisältäviä muotoja on alle prosentti kaikista aineiston yhdysosamuodoista. Muunlaisia yhdysosamuotoja on 2,6 % kaikista yhdysosamuodoista. Muiden muotojen ryhmään kuuluu muodostukseltaan vaihteleva joukko muotoja, jotka kuitenkin selvästi palautuvat

johonkin tiettyyn itsenäiseen sanaan. Ryhmään on laskettu esimerkiksi sellaiset itsenäisistä sanoista tyypistämällä muodostetut alkuosat, jotka eivät esiinny itsenäisinä sanoina.

Kirjallisuudessa useimmin muista erotettu tyyppi on *-nen*-loppuisten sanojen *-s*-loppuinen yhdyssosamuoto, joka vastaa niiden konsonanttivartaloa (esim. Vesikansa 1989: 225; ISK § 416; esim. *hampurilaispalaute*, *sijaisvaari*). Muodon yleisyyttä sekä uudessa että vanhassa yhdyssanastossa selittävät sen nominatiivia vastaavat merkitystehtävät sekä säännönmukainen muodostus suomen kielessä yleisistä *-nen*-loppuisista johdoksista. *-nen*-loppuinen sana muuttuu yhdyssanan alkuosana nominatiivia vastaavissa merkitystehtävissä toimiessaan lähes järjestelmällisesti *-s*-loppuiseen muotoon (*-nen*-loppuinen muoto on yhdyssanan alkuosassa mahdollinen vain nimistöissä). Tosin vierasta alkuperää olevia *-inen*-loppuisia sanoja vastaavissa merkityksissä käytetään myös *-i*-loppuista muotoa (ks. tarkemmin seuraava kappale). Myös eräistä kotoisista *-nen*-loppuisista adjektiiveista muodostetut *-s*-loppuiset yhdyssosamuodot tuntuvat mahdolltomilta (esim. *sulois-* ja *suolais-*). *-nen*-loppuinen sana voi esiintyä yhdyssanan alkuosana myös genetiivissä ja periaatteessa muissakin sijamuodoissa, mutta se on harvinaista. Tämän tutkimuksen aineistossa sekä yksikön että monikon genetiivejä on 6. Muita sijamuotoja ei esiinny aineistossa lainkaan. Yksikön genetiiveistä 4 on muodostettu sanoista, joista *-s*-loppuinen yhdyssosamuoto tuntuisi epäluontevalta (esim. *suloisenkömpelö*, *suolaisenhintainen*). Yhdessä *-s*-loppuinen muoto tekisi sanasta vaikean ymmärtää (*toisenvarainen*). Yhdessä sanassa myös *-s*-loppuinen muoto olisi mahdollinen, mutta genetiivialku tarkoittaa ilmauksen merkitystä (*hapettumisenestäjä*). Monikon genetiiveistä 4 on *nainen*-sanan esiintymiä, joissa *-s*-loppuinen muoto voisi luoda väärän mielikuvan yhdyssanan osien välisestä suhteesta (esim. *naistenhöylä*, ks. myös Tyysteri 2011c: 169–170). Kahdessa monikkotapauksessa myös *-s*-loppuinen muoto olisi luonteva (*maalaistenpaljastin* ja *pyöriäistentarkkailija*).

Vierasperäisiä *-inen*-loppuisia adjektiiveja vastaa yhdyssanan alussa usein *-i*-loppuinen muoto, joka ei vastaa niiden vartaloa (*mentaalitapaus*, *virtuaalihyödyke*, ks. myös ISK § 152, 268).⁷³ Tutkimusaineistossa yksikään pelkällä *-inen*-johtimella johdettu vierassana ei esiinny yhdyssanan alussa *-s*-loppuisessa yhdyssosamuodossa (sen sijaan *-minen*, *-mÄinen*- ja *-lÄinen*johdoksissa tämä on mahdollista, esim. *integroitumistrendi*, *nigerialaiskirje*). Oman kielitajuni mukaan *-(i)nen*-johdoksia vastaavat *-s*-loppuiset muodot ovatkin mahdollisia esisijaisesti summayhdyssanoissa

73 Tällaiset muodot on voitu saada kieleen joko takaperoisesti *-inen*-loppuisista sanoista johtamalla tai käännöslainatun yhdyssanan mukana suoraan yhdyssosamuodoksi. Esimerkiksi alkuosa *sosiaali*-on voinut saada alkunsa joko sanasta *sosiaalinen* (ensimmäinen muistiin merkitty esiintymä v. 1866 muodossa *sociallinen*) tai sanoista *sosiaalidemokraattinen* (ensimmäinen muistiin merkitty esiintymä v. 1879 muodossa *sosialdemokratillinen*) ja *sosiaalidemokratia* (ensimmäinen muistiin merkitty esiintymä v. 1889 muodossa *sosialidemokratia*, todennäköisesti käytetty myös aikaisemmin) (sanojen ensiesiintymistä ja etymologista ks. Koukkunen 1990 s.v. *sosiaalidemokraattinen*, *sosiaalidemokratia* ja *sosiaalinen*). Koska kaikilla *-i*-loppuisilla muodoilla on nykykielessä *-inen*-loppuinen pari, kutsun näitä muotoja tästä eteenpäin *-inen*-loppuisten sanojen yhdyssosamuodoiksi (en ainoastaan niitä vastaaviksi muodoiksi kuten edellä).

(esim. *sosiaalis-taloudellinen*), joita uudessa yhdyssanastossa on kaiken kaikkiaan vain vähän. Määrittelyyhdyssanoissa niiden käyttö tuntuu yleensä kielenvastaiselta tai vähintään vanhahtavalta (esim. *?seksuaaliskasvatus*, *?sosiaalistyö*).

Osa *-i*-loppuisista alkuosista esiintyy myös itsenäisenä lekseminä (esim. *automaatti*). Esimerkiksi yhdyssanoista *automaattimetro* ja *moraalikompassi* ei voida varmasti sanoa, onko niiden alkuosa *-inen*-loppuisen adjektiivin yhdyssamuoto (*automaattinen metro* > *automaattimetro*) vai onko yhdyssana muodostettu suoraan kahdesta substantiivista (*automaatti* + *metro* > *automaattimetro*). Koska rajanveto pelkän muodosteen merkityksen perusteella on monissa tapauksissa mahdotonta ja koska tässä tutkimuksessa yhdyssanojen taustalle ei oleteta semanttisia syvärakenteita, *-i*-loppuisiksi yhdyssamuodoiksi on luokiteltu vain ne alkuosat, jotka eivät esiinny itsenäisinä leksemeinä *-i*-loppuisessa muodossa. Rakenne on *-inen*-loppuisien vierasperäisten sanojen yhteydessä selvästi produktiivinen, vaikka sitä ei ole huomioitu aiemmassa tutkimuksessa (ISK § 268 mainitsee rakenteesta hyvin lyhyesti). Syynä rakenteen sivuuttamiseen lähinnä suomen kielen omiin ilmaisuvaroihin keskittyvässä sananmuodostuksen tutkimuksessa voi olla se, että rakenne esiintyy vain vierasta alkuperää olevien sanojen yhteydessä. Vierassanat ovat kuitenkin tärkeä ilmaisukeino, jota ei ole mitään syytä jättää sanastoa yleisesti käsittelevän saati uusien sanojen muodostusta koskevan tutkimuksen ulkopuolelle.

Myös joillain muilla kuin *-nen*-loppuisilla sanoilla on konsonanttivartaloa vastaava yhdyssamuoto. ISK:n (§ 416) mukaan ryhmään kuuluu muutamia *e*-vartaloisia nomineja sekä lisäksi sanat *enkeli* ja *jumala*⁷⁴. Kaikki ISK:n mainitsemat adjektiiviset vartalot (*nuor-*, *pien-*, *suur-*, *täys-* ja *uus-*) esiintyvät myös tämän tutkimuksen aineistossa. Substantiivivartaloista aineistossa esiintyy ainoastaan *pers-* (*perstasku* ja *perstuntuma*).⁷⁵ Vaikka substantiivisten konsonanttivartaloitten käyttö (lukuun ottamatta em. *-nen*-loppuisien sanojen vartaloita) ei ole yhtä selvästi produktiivista kuin adjektiivisten konsonanttivartaloitten käyttö, tähän ryhmään vanhastaan kuuluneiden sanojen mallin mukaan voidaan ainakin tietoisessa sananmuodostuksessa luoda myös uusia substantiivien konsonanttivartaloa muistuttavia yhdyssanan alkuosia. Näin on tehty ainakin tuotenimestä yleisnimitykseksi siirtyneessä tai siirtymässä olevassa sanassa *mölkkaari* 'kehikko, jonka takaa mölkkypelissä heitetään' (alkuosa ei vastaa *mölkky*-sanan konsonanttivartaloa, ja analyysissa se on luettu kategoriaan "muut").

Joitakin numeraaleja voi nominatiivin lisäksi edustaa yhdyssanan alkuosassa epäitsenäinen muoto (tarkemmin esim. ISK § 416). Tämän tutkimuksen aineistossa numeraalien yhdyssamuodoista esiintyvät lukusanojen *kolme* ja *neljä* *-i*-loppuiset yhdyssamuodot (esim. *kolmisäikeinen*, *nelikaistaistaa*) sekä lukusanojen *kaksi*, *viisi* ja *kuusi* muodot *kaks-*, *viis-* ja *kuus-* (esim. *kuuskakkoset* 'maastopuku M62'),

74 Vesikansa (1989: 221–222) pitää substantiivivartaloita loppuheittoisina muotoina ilmeisesti niiden runollisuuden vuoksi.

75 Kaikki ISK:ssa (§ 416) yhdyssanan alkuosana toimiviksi konsonanttivartaloiksi mainitut muodot eivät esiinny taivutuksessa varsinaisina konsonanttivartaloina. Tässä tutkimuksessa ne on kuitenkin laskettu ISK:n luokittelun mukaisesti konsonanttivartaloitten ryhmään.

jotka voidaan katsoa joko konsonanttivartalon kaltaisiksi muodoiksi tai puhekielenmukaisiksi loppuheittoisiksi muodoiksi (ks. myös Nikkilä 1994: 6–8; ISK § 770). Luvuista 1–9 on aineistossa nominatiivivariantti vain sanoista *yksi-*, *kaksi-* ja *viisi-*. Muodostetta *viisimauste* (myös muodossa *viismauste*) lukuun ottamatta kaikki aineiston nominatiivimuotoisella luvulla 1–9 alkavat yhdyssanat ovat *-inen*-adjektiivieja tai niiden johdoksia (esim. *viisiluukkuinen*, *kaksitarkoitteisuus*). Muilla lukusanoilla (esim. *nolla*, *tusina*, *sata*) ei ole erityistä yhdysosamuotoa, ja ne esiintyvät nominatiivissa myös muiden kuin *-inen*-adjektiivien alkuosina (esim. *nollapuhe* 'tyhjä puhe, tyhjät sanat', *tusinataulu* 'taulu, joka on samanlainen kuin hyvin monet muut').

Numeraalien yhdysosamuotoja on aineistossa 12 yhdyssanassa (numeraalialkuosista ks. myös luku 5.2.4). Lukusanojen lisäksi myös muutamilla värinnimityksillä on erityinen yhdysosamuoto. Värinnimitysten yhdysosamuodot eivät noudata yleisiä sananmuodostussääntöjä. Aineistossa värinnimen yhdysosamuoto on 97 alkuosassa. Edustettuna ovat kaikki ISK:ssa (§ 416) mainitut muodot: *kelta-*, *puna-*, *sini*, *valko-* ja *viher-*, joista *puna* ja *sini* tosin esiintyvät myös itsenäisinä sanoina. Yhdestäkään näistä ei esiinny aineistossa nominatiivivarianttia. Uudessa yhdyssanastossa yhdysosamuoto on siis näiden sanojen kohdalla käytössä yksinomaisena.

Yhdyssanan alkuosana voi toimia myös vokaalivartalo. Kirjallisuudessa usein mainittu tyyppi on joistakin *-A*-loppuisista substantiiveista vanhastaan käytetty *-o*-tai *-i*-loppuinen muoto (esim. *sikolätti*, *lehmisavu*) (Vesikansa 1989: 221–222; ISK § 416), joka käytännössä vastaa monikkovartaloa (esim. Hämäläinen 1883: 122–124). *o-* ja *i-*elementit on joskus tulkittu myös lisäainekseksi (ISK § 416). Monikkovartalon käyttöä yhdyssanamuotona on jo varhain pidetty murteellisuuutena (esim. Genetz 1881: 21–22; Hämäläinen 1883: 122–124), mutta se on ollut tapana esitellä yhdyssanojen muodostusta käsittelevissä kieliopissa ja muissa tutkimuksissa (esim. Vesikansa 1989: 221). ISK (§ 416) mainitsee tyyppin vanhahtavana. Tyyppi on hyvin harvinainen uusissa yhdyssanoissa, ja sitä käytetään vain sellaisista sanoista, joissa se on ollut käytössä jo vanhastaan. Tämän tutkimuksen aineistossa monikkovartalo on vain kahden yhdyssanan alkuosana (*sikopaja* ja *jalkorätti*). Substantiivin monikkovartaloiden lisäksi yhdyssanan alkuosana mahdollisia vokaalivartaloita ovat verbivartalot (esim. *leijukattila*, *riippupeppu*), joita valtaosa uuden yhdyssanaston alkuosina toimivista vokaalivartaloista on (aineistossa 32 kpl; ks. myös luku 5.3.2.1). (Yhdyssanan alkuosana toimivista verbivartaloista ks. myös ISK § 417.)

Verbit esiintyvät yhdyssanan alkuosassa vokaalivartalon lisäksi *-mA-* ja *-in-*loppuisissa muodoissa (*ampumapaini*, *kuolinapu*). Näistä *-mA*-loppuinen muoto vastaa päätteetöntä MA-infinitiiviä. *-mA*-alkuosa on uudessa yhdyssanastossa jonkin verran *-in*-loppuista alkuosaa suositumpi. Tätä selittävät *-mA*-alkuosan laajemmat käyttömahdollisuudet. Kuitenkin myös *-in*-alkuosa on pitänyt pintansa, vaikka jo Penttilä (1963 [1957]: § 223, s. 262–263) huomauttaa molempien tyyppien suosion vähenemisestä. ISK:n (§ 417) mukaan *-mA*-muotoa esiintyy sanakirja-aineistossa noin 60 eri alkuosasta, *-in*-muotoa taas paristakymmenestä. ISK:n (mp.) mukaan *-mA*-tyyppiin myös syntyy uusia alkuosia, *-in*-tyyppiin puolestaan ei.

Tämän tutkimuksen aineistossa *-mA*-loppu esiintyy 19 eri alkuosan kanssa, joista kaksi ei ole käytössä yhdessäkään KS:n yhdyshakusanassa (*kellumahiiri* ja *näkemä-este*). *-in*-loppu esiintyy 16 eri alkuosan kanssa, joista kolme ei ole käytössä yhdessäkään KS:n yhdyshakusanassa (*pakastinpuikko*, *poljinpinta*, *tehostinyksikkö*), mutta esiintyy kyllä itsenäisenä lekseiminä ja myös KS:n hakusanana (ks. myös Vesikansa 1989: 222; ISK mp.). Kontekstissaan näillä alkuosilla on kuitenkin enemmän verbaalinen kuin nominaalinen merkitys, mikä tukee niiden tulkitsemista uusiksi *-in*-muotoisiksi yhdyssanan alkuosiksi (esim. 22). Toisin kuin ISK (mp.) esittää, uusia *-in*-loppuisia yhdyssanan määriteosia saattaa edelleen syntyä lisää. Toisaalta esimerkki osoittaa havainnollisesti sen, että yhdyssanan alkuosan sanaluokka on monitulkintainen (ks. myös luvut 2.2.3 ja 5.2).

- (22) Tarkista, ettei [polkupyörän] polkimien *poljinpinta* ole kulunut sileäksi. Sileä pinta ei pidä, ja se on vaihdettava uuteen. [TT <- HS 3.4.2006, s. D1]

Vaikka sekä *-mA*- että *-in*-loppuiset alkuosat ovat uudessa yhdyssanastossa melko harvinaisia, niille ei ainakaan toistaiseksi ole kehittynyt vanhahtavaa tyyli-väriä, joka tekisi niistä uhanalaisia (vrt. edellä käsitellyt nominien monikkovartalot ja myöhemmin käsiteltävät *s*-linkintysolementin sisältävät muodot). Muodot esiintyvät myös uusien alkuosien yhteydessä ja näyttävät muutenkin saavan motivaationsa ennemmin rakenteen tarpeellisuuden kuin yksittäisten analogioiden kautta. Vokaalivartalon sekä *-mA*- ja *-in*-loppuisten alkuosien lisäksi verbaalisia merkityksiä voidaan ilmaista yhdyssanojen alkuosassa verbin *-minen*-nominaalistuksen *-mis*-loppuisella vartalolla tai muulla teonnimijohdoksella. Teonnimien nominaalistukset on tässä tutkimuksessa laskettu niiden muodon perusteella joko nominaatiivien tai *-nen*-loppuisen sanan *-s*-loppuisten alkuosien ryhmään. Nominaalisten teonnimijohdosten käyttö on uudessa yhdyssanastossa vokaalivartaloita sekä *-mA*- ja *-in*-loppuisia alkuosia suositumpi keino ilmaista yhdyssanan alkuosan verbaalisia merkityksiä, sillä pelkästään *-minen*-loppuisen teonnimijohdoksen *-mis*-vartalo on alkuosana 225:ssa aineiston yhdyssanassa ja *-ntA*-johtimella tuotettu teonnimijohdos 112:ssa aineiston yhdyssanassa (esim. *asumisperhe*, *kelluntahaalari*). Joskus erilaiset alkuosat esiintyvät rinnakkain samamerkityksisissä yhdyssanoissa (*uima-putkilo* ja *uintiputkilo*).

-s-loppuisia yhdysosamuotoja esiintyy myös muista kuin *-nen*-loppuisista sanoista (esim. *herrasmies*, *saarnastuoli*). Näissä sanoissa *s* voidaan tulkita eräänlaiseksi linkityselementiksi, sillä se ei periaatteessa kuulu sen enempää yhdyssanan alku- kuin jälkiosaankaan. Tyyppi on mahdollisesti saatu suomeen väärin segmentoitujen lainasanojen mukana (esim. ru *stol* sanassa *saarna|stuoli* > *saarnas|tuoli*), joista analoginen malli on levinnyt myös muihin yhdyssanoihin (Häkkinen 1990: 155). *s*-elementin käsittelyä alkuosan muodon yhteydessä tukee kuitenkin se, että se esiintyy vain tiettyjen alkuosien yhteydessä eikä jälkiosalla ole yhteyttä rakenteen esiintymiseen. ISK:n (§ 416) mukaan vakiintuneita *s*-elementin sisältäviä yhdysosamuotoja on parisenkymmentä. Tämän tutkimuksen aineistossa rakenne esiintyy 10:n eri alkuosan kanssa (12 eri yhdyssanassa). Mahdollisuutta muodostaa

analogian kautta uusia tätä rakennetta noudattavia alkuosia ei voida sulkea pois, mutta rakenne näyttää vakiintuneen vain sellaisten sanojen ominaisuudeksi, jotka esiintyvät tässä muodossa myös jossakin vanhassa ja vakiintuneessa sanassa (esim. *herraskeppi*, *rouvasporno*). Samat sanat voivat esiintyä yhdyssanan alkuosana myös ilman *s*-linkityselementtiä (esim. *herrakertymä* 'rikkaille jo varhain kertyvä eläke', *rouvamissi*).

Tutkimusaineiston yhdysosamuotoisista alkuosista 2,6 % on luokiteltu kategoriaan "muut". Tähän ryhmään kuuluvat elementit ovat itsenäistä sanaa edustavia epäitsenäisiä muotoja, jotka eivät kuitenkaan noudata mitään johdonmukaista sananmuodostusperiaatetta. Tyypillisiä ryhmän jäseniä ovat esimerkiksi tyypistämällä muodostetut alkuosat, jotka eivät kuitenkaan esiinny itsenäisinä leksemeinä (esim. *aluvanne* 'alumiinivanne', *innoagentti* 'innovaatioagentti'). Muut-kategoriaan kuuluvat myös alkuosat, joiden muodostuksessa on todennäköisesti pyritty saamaan aikaan rakenteellinen samankaltaisuus jonkin olemassa olevan rakennetyypin kanssa siinä kuitenkaan aivan täysin onnistumatta (esim. em. *mölkkaari*). Muut-kategorian alkuosat tarjoavat mielenkiintoisen tutkimuskohteen esimerkiksi yksittäisten sanojen motivaatiosta ja analogiasta kiinnostuneille, mutta tämän tutkimuksen piiriin niiden yksityiskohtainen analyysi ei kuulu.

5.3.2.4 Genetiivit

Tutkimusaineiston yhdyssanoista 3,9 % (865 kpl) on genetiivialkuisia. Genetiivi mainitaan suomen yhdyssanoja käsittelevissä tutkimuksissa yleensä nominatiivimuotoisen alkuosan jälkeen käytetympänä tai jollakin tapaa implisiittisesti toiseksi tärkeimpänä alkuosavaihtoehtona. Uudessa yhdyssanastossa epäitsenäiset leksikaaliset elementit ovat kuitenkin genetiiviä yleisempiä – sekä kotoisia että vierasperäisiä prefiksinkaltaisia elementtejä on tutkimusaineistossa enemmän kuin genetiivimuotoisia alkuosia (genetiivialkuosia on 3,9 % koko tutkimusaineistosta, kun kotoisten ja vierasperäisten prefiksinkaltaisten alkuosien osuudet ovat 4,5 % ja 4,2 % aineistosta). Genetiivin kanssa lähes yhtä yleinen on *-nen*-loppuisten sanojen *-s*-loppuinen yhdysosamuoto (*-s*-loppuisten yhdysosamuotojen osuus 3,5 % aineistosta).

Uusissa yhdyssanoissa nominatiivialkuja esiintyy myös sellaisissa sanoissa, joissa genetiivi tuntuisi odotuksenmukaisemmalta (esim. *islampelko*, *jalkalämmittin*, *kanaihoinen*). Nominatiivialkuosan valinta on kielenkäyttäjälle keino osoittaa kokonaisuuden käsitteellistä yhteenkuuluvuutta, sillä toisin kuin genetiivialkuisen muodoste, nominatiivialku osoittaa suomen kielessä rakenteen yksiselitteisesti yhdyssanaksi. Nominatiivimuodon valintaan voi johtaa myös epävarmuus genetiivialkuisen muodosteiden oikeinkirjoituksesta.

Tiedot genetiivikategorian sisäisestä jakaumasta on koottu taulukkoon 10.

TAULUKKO 10. Tutkimusaineiston yhdyssanojen genetiivialkuosien muotojakauma ($n = 865$)

Alkuosan muoto	Esimerkki	Osuus genetiivi-alkuisista	Osuus kaikista alkuosista	Lukumäärä
Yksikkö	<i>japaninsinappi, puheentunnistus</i>	95,5 %	3,7 %	826
Monikko	<i>lastenkännykkä, ripsientaivutin</i>	3,6 %	0,1 %	31
Monikko (<i>in</i> -päätte)	<i>harvainyhtiö, vanhempainkoulu</i>	0,8 %	0,03 %	7
Possessiivisuffiksillinen muoto	<i>itsensäpaljastamo</i>	0,1 %	0,005 %	1
Yhteensä		~ 100 %	~ 3,9 %	865

Genetiivimuotoisista yhdyssanan alkuosista uudessa yhdyssanastossa selvästi yleisin on yksikkö (95,5 % kaikista aineiston genetiivialuista). Myös monikon genetiivi on mahdollinen yhdyssanan alkuosan muoto, mutta huomattavasti yksikköä harvinaisempi (kaikki monikon genetiivin muodot yht. 4,4 % kaikista genetiivialuista). Yhdyssanojen alkuosissa esiintyvät kaikki nykykielessä tavattavat monikon genetiivin päätteet: sekä monikkovartaloon liittyvät päätteet *-den* ~ *-tten* (*maiden* ~ *maitten*) ja *-en* (*vanhempien, pienien*) että yksikkövartaloon liittyvät päätteet *-ten* (*pienten*) ja *-in-* (esim. *vanhempain-*) (monikon genetiivin päätteistä tarkemmin esim. ISK § 83). Tyyliltään muista eroaa yksikkövartaloon liittyvä vanhahtava *-in-*päätte. Koska monikon genetiivi on kaiken kaikkiaan uusien yhdyssanojen alkuosana melko harvainen, olen erottanut luokittelussa omaksi kategoriakseen vain tämän tyyliltään selvästi muista eroavan päätteiden (0,8 % kaikista genetiivialuista). Uusien yhdyssanojen alkuosissa *-in-*päätte esiintyy vain sellaisten sanojen yhteydessä, joihin muoto on vanhastaan vakiintunut. Rakennetta sinänsä ei siis voi pitää produktiivisena, vaan rakennetta noudattavia uusia sanoja tuottavat tietyt tässä muodossa produktiiviset lekseemit (aineiston sanoissa *harvain, kansain, vanhempain* ja *varain* sekä sanassa *taalainkoivu* esiintyvä alueeseen viittaava *taalain* 'alku- jaan Taalainmaalta löydetty koivulaji').

Jussi Niemi (2009: 252) esittelee CD-PS:n yhdyssanojen alkuosan muotojakaumaa koko sanaluokan mukaan jaoteltuna (yhdyssubstantiivit, -adjektiivit ja -verbit). Taulukkoon 11 on koottu tiedot yksikön genetiivin esiintymisestä eri sanaluokkiin kuuluvissa kaksiosaisissa yhdyssanoissa. Niemen (mp.) rajauksen mukaisesti vertailusta on jätetty kokonaan pois alkuosana toimivat muotosanat.

TAULUKKO 11. Genetiivialkuisuus koko muodosteen sanaluokan mukaan tutkimusaineistossa ja CD-PS:ssä

	TT ($n = 20\ 827$)	CD-PS (Jussi Niemi 2009: 252, $n = 2\ 697$)
Adjektiivit	9,4 %	18,1 %
Substantiivit	3,8 %	11,3 %
Verbit	3,4 %	11,8 %

Jussi Niemi (2009: 252) käyttää tutkimuksessaan hyvin erilaisia alkuosan muotokategorioita kuin tässä tutkimuksessa on käytetty (Niemen kategoriat ovat perussana, johdos, yksikön genetiivi, muu sijamuoto ja taivutettu johdos), joten erojen tilastollinen testaus ei ole mielekästä. Taulukossa 11 olevat luvut kuitenkin osoittavat selvästi, että genetiivialkuisuus on uudessa yhdyssanastossa harvinaisempaa kuin CD-PS:ssa. Ero on selvä sekä yhdysadjektiivien, -substantiivien että -verbien ryhmässä.

5.3.2.5 Taipumattomat ja vaillinaisesti taipuvat sanat

Tutkimusaineiston sanoista 1 %:n (241 kpl:n) alkuosana toimii taipumaton tai vaillinaisesti taipuva sana. Ryhmään kuuluu adverbejä ja adpositioita (*hyvinpitely, kanssakädellinen, läpivuoto ympärileikkaaja*), kiteytyneitä taivutusmuotoja (*perässähiihtäjä, päältäkastelu*), taipumattomia adjektiiveja (*ensineule, pikkueuro*), keskustelupartikkeleja (*anteeksikone, juupaseipästely*) sekä muutamia interjektioita ja onomatopoeettisia ilmauksia (*wau-arkkitehtuuri, plim-pitsa, vuhjooga*). Lisäksi mukana on yksi seitsemässä sanassa toistuva taipumaton substantiivi (*syysliukkaat*) sekä kahdessa sanassa toistuva taipumaton indefiniittipronomini (*jokanuori*).

Taipumattomat nomininit olisi periaatteessa mahdollista luokitella myös nominatiivimuotoisiksi, mutta niiden lukeminen taipumattomien sanojen ryhmään kuvastaa paremmin niiden poikkeuksellista luonnetta yhdyssanojen muodostuksessa. Taipumattomien sanojen kohdalla yhdyssanaa muodostavan kielenkäyttäjän ei tarvitse tehdä valintaa eri muotojen välillä, sillä mahdollisia muotoja on vain yksi. Sen sijaan monien taipumattomien sanojen yhteydessä kielenkäyttäjä joutuu pohtimaan yhteen ja erikseen kirjoittamista erityisen tarkkaan, kirjoitetaanhan taipumattomat adjektiivit oikeinkirjoituksen pääsäännön mukaan erilleen pääsanastaan, elleivät sanat yhdessä muodosta erityistä käsitettä. Myös monien vajaan taipuvien sanojen oikean kirjoitusasun valinta vaatii kiteytyneisyyden tarkkaa pohdintaa (vrt. *perään ajaja* vai *peräänajaja*; ks. myös T. Kolehmainen 2014: 165–175).

190 kpl (78,8 % tämän alakategorian sanoista ja 0,9 % kaikista alkuosista) on kokonaan taipumattomia ja 51 kpl (21,2 % tämän alakategorian sanoista ja 0,2 % kaikista) vajaan taipuvia. Vajaan taipuvilla sanoilla on kaksi tai useampia taivutusmuotoja (ei kuitenkaan täyttä taivutusparadigmaa⁷⁶), jotka voivat olla joko nykyistä nominien sijataivutusta vastaavilla päätteillä muodostettuja tai aiemmissa kielen vaiheissa käytettyihin sijamuotoihin kivettyneitä muotoja (vrt. *altatuulettuvasti, ristiinomistus* ja *alaskirjata, kotonakulkija*). Koska vain harvoilla vajaan taipuvilla sanoilla on varsinainen nominatiivimuoto, vajaan taipuvien sanojen eri muodot eivät yleensä ole yhdyssanoissa keskenään vaihtoehtoisia (esim. *perässähiihtäjä* ja *peräänajaja*). Tosin joissain tapauksissa nominatiivialkuisten yhdysnominien avointa osienvälistä merkityssuhdetta vastaava rakenne on mahdollista luoda käyttämällä alkuosana pelkkää vajaan taipuvan sanan vartaloa (*harhaheitto* ja *harhaanheitto*).

⁷⁶ Esim. Koiviston (2013: 40–41) mukaan kaikki vajaan taipuvan paradigman muodot tulisi katsoa omiksi lekseemeikseen.

5.3.2.6 Muut sijamuodot

ISK (§ 418–419) mainitsee nominatiivin ja genetiivin ohella yleisimmiksi alkuosan sijoiksi paikallissijat, mutta huomauttaa myös instruktiivin ja abessiivin olevan mahdollisia. Lisäksi muodosteen jälkiosan ollessa partisiippi tai sen johdos, alkuosa voi olla myös partitiivissa. Vesikansa (1989: 221) mainitsee yhdyssanoissa nominatiivin ja genetiivin lisäksi mahdollisina sijoina vain paikallissijat.

Muut sijamuodot kuin nominatiivi ja genetiivit ovat uuden yhdyssanaston alkuosissa hyvin harvinaisia (vain 28 kpl eli 0,1 % tutkimusaineiston alkuosista). Tiedot tutkimusaineiston muiden sijojen sijajakaumasta on koottu taulukkoon 12. Esiintymien vähyden vuoksi taulukossa on ilmoitettu ainoastaan absoluuttiset määrät.

TAULUKKO 12. Tutkimusaineiston yhdyssanojen muussa sijassa kuin nominatiivissa ja genetiivissä olevien alkuosien muotojakauma ($n = 28$)

Alkuosan muoto	Esimerkki	Lukumäärä	
		yks.	mon.
Inessiivi	<i>rännässätallaaja, työssäoppija</i>	5	0
Elatiivi	–	0	0
Illatiivi	<i>maahanimeyttämö, lapsinsekaantuja</i>	12	1
Adessiivi	<i>hällävälästi, nukeillanäyttelijä</i>	1	1
Ablatiivi	–	0	0
Allatiivi	<i>maallemuuttaja, maallemuutto</i>	2	0
Instruktiiv	<i>käsinlyöty, moninpeli</i>		2
Abessiivi	–	0	0
Partitiivi	<i>vettähylykivä</i>	1	0
Translatiivi	<i>puheeksiotto, puhtaaksiviljelty</i>	3	0

Kuten taulukosta 12 nähdään, muista sijoista yleisin on yksikön illatiivi (12 kpl). Elatiivia ja ablatiivia lukuun ottamatta kaikki muut paikallissijat esiintyvät aineistossa vähintään kerran, mutta monikkomuotoja on vain illatiivista ja adessiivista (1 kpl kummastakin). Muista ISK:n (§ 418–419) mainitsemista mahdollisista sijoista aineistossa esiintyvät myös partitiivi ja instruktiivi, mutta abessiivimuotoisia alkuosia tutkimusaineistossa ei ole yhtäkään. ISK:ssa (mp.) mainittujen sijojen lisäksi tutkimusaineistossa on 3 translatiivisijaisen alkuosan sisältävää yhdyssanaa.

Kielioppien, sanaston yleiskuvausten ja aiempien yhdyssanoja koskevien tutkimusten toteamukset siitä, että nominatiivin ja genetiivin lisäksi myös muut sijamuodot ovat mahdollisia suomen kielen yhdyssanojen alkuosassa, pitävät tietysti kieliopin näkökulmasta paikkansa. Kielen käytön näkökulmasta muut sijamuodot ovat kuitenkin uuden yhdyssanaston alkuosissa marginaalisessa asemassa. Rakennetta noudattavat uudet yhdyssanat syntyvät usein sanaliitosta tiivistymisen kautta (esim. 23; käsittelen uusien yhdyssanojen muodostusta tarkemmin luvussa 6.1).

- (23) Potilan tiilitehdas Hämeenkyrössä valmistaa monen sortin tiiliä. Korjausrakentajalle paikka on oikea apaja, josta eläväpintaisen kakkoslaadun lisäksi voi tilata harvinaisia *käsinlyötyjä* tiiliä. [TT <- Meidän Talo, 11/2004, s. 84]

Toisessa kontekstissa sanaliiton käsitteellinen yhteenkuuluvuus ei välttämättä olisi yhtä selvää, ja rakenne voitaisiin helposti ymmärtää myös sanaliitoksi. Monia muussa sijassa olevan alkuosan sisältäviä yhdyssanoja voidaankin pitää syntaktista johtamista vastaavan kontekstisidonnaisen yhdistämisprosessin kautta syntyneinä (synkronisesta johtamisesta ks. tarkemmin esim. Koski 1978: 103). Tällaisista sanoista todennäköisesti vain harvat vakiintuvat leksikkoon.

ISK:n (§ 403, 418) mukaan muut sijat esiintyvät alkuosassa yleensä deverbaalin jälkiosan kanssa. Omassa tutkimusaineistossani on kaksi sellaista pronominalkuista yhdyssanaa, joiden alkuosa on muussa sijamuodossa kuin nominatiivissa tai genetiivissä mutta joiden jälkiosa ei ole deverbaali (esim. 24 ja 25).

- (24) Onko nihilisti vain *hällävälisi*? [otsikko] – ”Nuo nihilistit hyökkäävät nihilismin kimppeun raivolla, jota ei toisilta filosofeilta löydy”, kirjoittaa valtiotieteiden tohtori Kalle Haatanen (s. 1969). – Nihilistin hän suomentaa *hälläväliseksi*. [TT <- HS 28.11.2008, Johan Lahdenperä, s. C 3]
- (25) Pelaamaan pääsee eri osapuolilla ja lisäksi tarjolla on yksittäisiä skenaarioita. Myös *moninpelin* karttoja voi pelata tietokonetta vastaan. [TT <- Metro, 8.6.2006, Kim Soares, s. 21]

Yhdyssana *hällävälisi* on johdettu sanaliitosta *hällä väliä* (vrt. *hän*-sanan merkitykset SMS s. v. *hän*). Yhdyssanan alkuosa on adessiivissa kuten sanaliitossakin. Yhdyssanan poikkeuksellinen rakenne selittyy siis sanaliitosta johtamisella. *Moninpeli*-sanan muodostukseen on todennäköisesti otettu mallia *kaksinpeli*- ja *nelinpeli*-sanoista, joilta se on perinyt myös alkuosansa muodon. *Moninpeli*-sanan alkuosa voidaan ymmärtää paitsi instruktiivisijaiseksi pronominiksi myös adverbiksi (esim. Penttilä tulkitsee *monin*-alkuosan instruktiiviksi esimerkiksi *moninkertainen*-yhdyssanassa, 1963 [1957]: § 223, s. 270–271). Instruktiivia voidaan pitää eräänlaisena adverbistimenä, sillä instruktiivissa olevat substantiivit saavat usein substantiivin määritteiden sijaan adverbien määritteitä tai esiintyvät osana adverbimäistä kiteymää (ISK § 686, 1261). Esimerkiksi sellaiset sanastoon vakiintuneet instruktiivisijaisen alkuosan ja ei-deverbaalin jälkiosan sisältävät yhdyssanat kuin *moninkertainen* ja *kaikinpuolinen* lienevät syntyneet juuri tällaisten kiteymien kautta (*monin kerroin* > *moninkertainen*, *kaikin puolin* > *kaikinpuolinen*). Tällaisissa kiteymissä esiintyvät tai muuten kantasanansa merkityksestä eriytyneet mutta niiden instruktiivimuodon kanssa identtiset muodot voidaankin lukea myös adverbeiksi (ISK § 388). Adverbialkuosien ja ei-deverbaalien jälkiosien yhdistyminen yhdyssanaksi ei ole samalla tavalla rajoitettua kuin eri sijamuodoissa olevien nominaalukuosien ja ei-deverbaalien jälkiosien, joten *moninpeli*-muodosteen poikkeuksellinen rakenne voidaan selittää alkuosan adverbien ja muodosteen motivaationa olevan analogian kautta. Myös tätä rakennetta noudattava yhdyssanat ovat uudessa sanastossa melko harvinaisia (*moninpelin* lisäksi esim. *nurinpäinen*).

5.3.3 Yhdyssanan osien kongruenssi

Yhdyssanan saamat kieliopilliset tunnukset liittyvät yleensä vain muodosteen jälkimmäiseen osaan. Kuitenkin joissain adjektiivialkuisissa yhdyssubstantiiveissa (*vanhapoika* : *vanhanpojan*, *uusivuosi* : *uudenvuoden*) ja yhdysnumeraaleissa (*yhdeksänkymmentä* : *yhdeksääkymmentä*) sekä kahdesta pronominivartalosta syntyneissä pronomineissa *joku* ja *jompikumpi* taivutustunnukset liittyvät sekä alku- että jälkiosaan. (NES s. v. *joku*; ISK § 103, § 404, § 420, § 780.) ISK:n (§ 404, § 420) tapaan kutsun tällaisia yhdyssanoja yhdysosiltaan kongruoiviksi. Taivutustunnusten lisäksi kieliopillisista tunnuksista yhdyssanan alkuosaan voivat tietyissä tapauksissa liittyä myös komparatiivin tunnus sekä liitteet *-kin* ja *-kAAAn* (*laaja-alaisempi* ~ *laajempialainen*, *muunlainenkin* ~ *muunkinlainen*) (ISK § 404, § 428). Sekä taivutustunnusten että muiden kieliopillisten elementtien tunkeutuminen yhdyssanan sisään on merkki siitä, että muodoste ei ole morfologisesti täysin koskematon ja eheä (ks. myös luku 2.1.3). Tässä tutkimuksessa tarkastelen ainoastaan yhdysosien välistä kongruenssia, en muiden kieliopillisten tunnusten liittymistä yhdyssanaan.

Perinteisen kieliopillis-kielenhuollollisen näkökulman mukaan yhdyssanan osien kongruenssi tai sen puuttuminen nähdään yhdyssanan itsensä ominaisuutena. Yhdyssanat voidaan jakaa a) osiltaan kongruoimattomiin, b) osiltaan kongruoiviin sekä c) osiltaan valinnaisesti kongruoiviin. Vaikka kongruenssi on perinteisesti nähty yhdyssanaan kiinteästi kuuluvana ominaisuutena, aiemmissä tutkimuksissa viitataan kongruenssiin myös kielenkäyttäjän valintana varsinkin niissä tapauksissa, joissa sitä pidetään kieliopillisesta ja joskus myös kielenhuollollisesta näkökulmasta valinnaisena. (R. Lehtinen 1984; Eronen 1996; T. Kolehmainen 2014: 157–165.)

Uusien yhdyssanojen kongruenssin tutkiminen kielen käytön näkökulmasta on sikäli ongelmallista, että niiden frekvenssi voi olla esimerkiksi internetaineistossa tai korpuksessa hyvin pieni. Hakukonehauissa osumia ei välttämättä saada lainkaan. Vaikka osumia tulisi, tutkimuksessa olisi pystyttävä määrittelemään kynnyksarvo sille, milloin sanan voidaan ajatella olevan ainakin joidenkin kielenkäyttäjien mielessä osiltaan valinnaisesti kongruoiva. Yksi tai kaksi osumaa ei välttämättä riitä todisteeksi aidosta variaatiosta. Tämän vuoksi käytän tarkasteluni pohjana kieliopillis-kielenhuollollista näkökulmaa ja sen kolmea edellä mainittua kategoriaa. Perustan sanojen luokittelun pääosin sanakirjatietoon ja tietoon kielenhuollon tekemistä ratkaisuihin sekä omaan kieli-intuitiooni. Mahdollisuuksien mukaan havainnollistan kongruenssin valinnaisuutta internetistä kerättyjen frekvenssitietojen avulla (hakukonehauista saatuja frekvenssitietoja voidaan pitää vain suuntaa-antavina; hakukonehauista tutkimusmenetelmänä esim. Petrova 2011: 27–34; Heinonen 2013: 26–29).

Tutkimusaineistossani ei ole yhtään osiltaan säännönmukaisesti aina kongruoivaa yhdyssanaa. Sen sijaan kieliopillis-kielenhuollollisesta näkökulmasta osiltaan valinnaisesti kongruoivia yhdyssanoja on aineistossa kuusi. Tapausten harvinaisuuden vuoksi poikkean tässä luvussa tutkimuksen yleisestä linjasta ja käsittelen aineistoani pääasiassa kvalitatiivisesti. Koska tavalliset kielenkäyttäjät eivät aina

jaa adjektiivialkuisia yhdyssubstantiiveja kongruoiviin, valinnaisesti kongruoiviin ja kokonaan kongruoimattomiin yhtä tarkasti kuin tutkijat ja kielenhuoltajat, pelkää kielenkäytön näkökulmasta tarkasteltuna osiltaan valinnaisesti kongruoivia tapauksia olisi enemmän. Esimerkiksi Raija Lehtisen (1984: 517; ks. myös Kolehmainen 2014: 159–165) mukaan adjektiivit *vapaa*, *hapan*, *raaka* ja *vajaa* eivät yhdyssanan alkuosana toimiessaan koskaan kongruoi jälkiosan kanssa. Näin ollen esimerkiksi yleiskieleen lainautuneena laskuvarjohyppyterminä tutkimusaineistooni päätyneen *vapaapudotus*-sanan osien ei tulisi kongruoida edes valinnaisesti (em. kieliopillis-kielenhuollollisen näkökulman mukaisesti olen näin ollen analysoinut sanan osiltaan kongruoimattomaksi). Oman aineistoni konteksteissa sana esiintyy vain nominatiivissa, mutta 12.2.2013 tekemiäni Google-hakujen mukaan *vapaatapudotusta* käytetään internetissä laajasti myös osiltaan kongruoivana (esim. kongruoiva *vapaanpudotuksen* 1 200 osumaa; vrt. kongruoimaton *vapaapudotuksen* 9 680 osumaa).

Aineistoni kuudesta osiltaan valinnaisesti kongruoivasta tapauksesta kaksi (*mustapekka* ja *omatunto*) on vanhoja sanoja, jotka ovat päätyneet aineistoon uudenlaisen käyttötapsansa vuoksi. Vanhastaan käytössä olleen ilmauksen kanssa homonyminen *isojako* puolestaan esiintyy aineistossa uudessa merkityksessä 'yökerhossa ja sen edustalla valomerkin jälkeen tapahtuva pariutuminen'.⁷⁷ Aineiston kuudesta osiltaan valinnaisesti kongruoivasta sanasta kolme (*mustasuola*, *pitkäriisi* ja *puoliviikko*) on todennäköisesti aitoja morfologis-semanttisia uudismuodosteita (uudissanojen jaosta morfologis-semanttisiin ja semanttisiin ks. esim. L. Hakulinen 1979: 381–425).

Omatunto esiintyy aineistossa uuden termimäisen liittosanan osana (esim. 26). Sana on kontekstissaan genetiivimuotoisena ja myös sen alkuosa *oma* on taipunut. Myös KS 2012 (s. v. *omatunto*) selittää sanan aina osiltaan kongruoivaksi.

- (26) Merkittäväksi megatrendiksi Vanjoki arvioi tällä hetkellä *globaalien omantunnon*. – –. ”Nyt on nousussa yleinen ymmärrys siitä että kuulumme johonkin kokonaisuuteen. Tämä kokonaisuus käsittää maapallon.” [TT <- Talouselämä 43, 14.12.2001, Mirva Heiskanen, s. 16]

Omatunto-yhdyssanan osien kongruenssin katoamista on uumoiltu jo joitakin vuosikymmeniä (esim. R. Lehtinen 1984), eivätkä pelkät tutkimusaineiston kontekstitiedot ja sanakirjamääritelmä riitä osoittamaan sanaa aina osiltaan kongruoivaksi. Taulukkoon 13 on koottu *omatunto*-yhdyssanan käyttötapojen määriä 1.2.2013 tehdystä Google-hausta. Yksikön ja monikon kieliopillisten sijojen lisäksi taulukossa on esitetty tiedot yksikön sisä- ja ulkopaikallissijoista sekä monikon allatiivista, adessiivista ja instruktiivista. Monikon sijoista juuri instruktiivi on otettu mukaan siksi, että *omatunto*-yhdyssana esiintyy monikossa yleisesti juuri tässä sijamuodossa (esim. vakiintuneessa sanonnassa *puhtain omintunnonin*). Vakiintuneet

⁷⁷ Uudella tavalla käytettyjen vanhojen tai vanhojen ilmausten kanssa homonymisten yhdyssanojen määrä on osiltaan kongruoivien sanojen pienessä joukossa selvästi suurempi kuin aineistossa yleisesti (ks. myös luku 4.1).

sanonnat (esim. *hyvällä omallatunnolla*) voivat lisätä osiltaan kongruoivan variantin suosiota myös muissa sijoissa. Possessiivisuffiksillisista muodoista ovat näytteenomaisesti mukana kielipillisten sijojen yksikön ensimmäisen persoonan muodot. Yleisin käytötapa on lihavoitu kullakin rivillä.

TAULUKKO 13. Yhdyssien kongruenssi *omatunto*-sanassa Google-hakujen (2/2013) mukaan.

Koko muodosteen sija ja luku	Osat kongruoivat (esim. <i>omantunnon</i>)	Osat eivät kongruoi (esim. <i>omatunnon</i>)	Osat kongruoivat, kirjoitettu erikseen (esim. <i>oman tunnon</i>)	Osat eivät kongruoi, kirjoitettu erikseen (esim. <i>oma tunnon</i>)
Yks. gen.	202 000	27 000	17 400	101
Yks. part.	133 000	51 200	33 900	921
Yks. gen. + yks. 1. pers. poss. suff.	28 000	72 600	269	312
Yks. part. + yks. 1. pers. poss. suff.	23 600	5 910	818	6
Yks. ess.	8 070	4 720	124	2
Yks. transl.	6 050	2 660	1 160	0
Yks. iness.	15 000	6 140	1 400	9
Yks. elat.	36 000	10 300	4 100	9
Yks. illat.	16 600	14 800	1 240	5
Yks. allat.	13 300	33 100	1 490	5
Yks. ablat.	7 180	1 280	126	0
Yks. adess.	125 000	32 700	51 900	394
Mon. gen.	1260	380	(4 950)	0
Mon. part.	4 070	2 880	(3 710)	0
Mon. all.	454	293	(122)	0
Mon. adess.	245	83	(144)	0
Mon. instr.	3 100	2 280	358	1

Kuten taulukosta 13 nähdään, osien kongruenssi on yksikön allatiivia (*omalletunnolle* ~ *omatunnolle*) ja yksikön possessiivisuffiksillisista genetiiviä (*omantuntoni* ~ *omatuntoni*) lukuun ottamatta kaikissa esitellyissä sijoissa yleisempää kuin kongruenssin puuttuminen. Eron suuruus kuitenkin vaihtelee sijamuodoittain. Oikeinkirjoitussääntöjen mukainen yhteen kirjoittaminen on kaikissa sijoissa yleisempää kuin norminvastaisuus. Huomiota herättää yksikön adessiivi, jossa kongruenssi ja erikseen kirjoittaminen (*omalla tunnolla*) on yleisempää kuin kongruenssin puuttuminen ja yhteen kirjoittaminen (*omatunnolla*; alleviivattu). Osiltaan kongruoivat yhdyssanat voivat vaikeuttaa yhdyssanojen oikeinkirjoitussääntöjen tulkittamista entisestään, joten *omalla tunnolla* -variantin käyttö voi olla seurausta oikeinkirjoitusvaikeuksista yhdistettynä haluun käyttää edellä mainittuun *hyvällä omallatunnolla* -sanontaan vakiintunutta kongruoivaa varianttia. Osien kongruenssi ja erikseen kirjoittaminen näyttää olevan osien kongruenssia ja yhteen kirjoittamista yleisempää myös monikon genetiivissä, partitiivissa ja adessiivissa. Näissä

sijoissa on kuitenkin huomattava, että tutkittava ilmaus on homonyyminen yleensä juuri monikossa käytettävän nominista ja sen määritteenä olevasta adjektiivista muodostuvan *omat tunnot* -sanaliiton kanssa (esim. – – *tärkeää on sijaisvanhemmille tarjottava mahdollisuus omien tuntojen jakamiseen*).

Omantunnon lisäksi sanat *isojako* ja *puoliviikko* esiintyvät aineistossani kontekstissa, jossa niiden osien kongruenssi näkyy (esimerkit 27 ja 28).

- (27) Yökerhossa saan säälikatseita: hymy ja päänkallistus ilmaisevat, että ”kyllä mä ymmärrän”. *Isonjaon* aikaan epätoivoisimmat kysyvät, pystynkö vällyjen välissä mihinkään. [TT <- Anna 25.9.2001, s. 66]
- (28) Mahatautipöpön alkuperä varmistuu *puolenviikon* maissa [TT <- Aamulehti 28.3.2000, s. A7]

Uudenlaisessa merkityksessä esiintyvän *isojako*-sanana osien kongruenssi juontaa todennäköisesti sen mallina toimineesta ja sen kanssa homonyymisestä *isojako*-sanasta ’1700-luvun maanjako’, jonka osat voivat kongruoida keskenään (KS s. v. *isojako*). *Puoliviikko* puolestaan on voinut saada mallin osien kongruenssiin *puolipäivä*-sanasta, jossa sekä kongruenssi että kongruoimattomuus ovat KS:n (s. v. *puolipäivä*) mukaan mahdollisia. 12.2.2013 tekemieni Google-hakujen perusteella *puoliviikko*-sanana osien kongruenssi on jonkin verran yleisempää kuin osien kongruoimattomuus.

Sanat *mustasuola*, *mustapekka* ja *pitkäriisi* esiintyvät aineistossa nominatiivimuotoisina, joten niiden osien kongruenssista ei voida esimerkikikäyttöjen perusteella päätellä mitään. *Mustapekka* on KS:n (s. v. *mustapekka*) mukaan osiltaan kongruoiva, ja sekä sitä että *mustasuolaa* ja *pitkäriisiä* käytetään 12.2.2013 tekemieni Google-hakujen mukaan sekä osiltaan kongruoivina että kongruoimattomina. Sanakirjatiedon, internetosumien ja kieli-intuition lisäksi näiden sanojen tulkitsemista valinnaisesti kongruoiviksi tukee myös se, että a) niiden alkuosana on tyypillisesti kongruoivana esiintyvä adjektiivi ja b) ne ovat sanaliitoista (*mustapekka* sanaliittomuotoisesta erisnimestä) tiivistyneitä.

Edellä olen tarkastellut omassa tutkimusaineistoni yhdyssanojen osien kongruenssia. Seuraavaksi käsittelen osiltaan kongruoivien yhdyssanojen syntytapoja ylipäänsä. Tyypillisiä jälkiosan kanssa kongruoivia alkuosia ovat esimerkiksi sellaiset adjektiivit kuin *iso*, *oma*, *musta*, *nuori* ja *pitkä*. Vaikka suurin osa alkuosana kongruoivista adjektiiveista kuuluu varsin vakiintuneeseen sanastoon, myös uudet adjektiivit voivat sanaliiton kautta yhdyssanaan päätyessään edelleen kongruoida muodosteen jälkiosan kanssa. Yksikään adjektiivi ei kongruoi jälkiosan kanssa aina yhdyssanan alkuosana toimiessaan, vaan sana voi olla toisissa muodosteissa kongruoiva (*omallatunnolla*) ja toisissa kongruoimaton (*omalääkärillä*). (ISK § 420.) Mitään yksiselitteistä sääntöä sille, milloin yhdyssanan osat kongruoivat ja milloin eivät, ei ole olemassa (Eronen 1996: 12, ks. myös R. Lehtinen 1984: 517).

Osiltaan kongruoivan yhdyssanan synty on teoreettisesti todennäköisintä silloin, kun yhdyssana syntyy sanaliitosta tiivistymällä (*musta vyö* : *mustan vyön* >

mustavyö : *mustanvyön*, esimerkki ei TT:stä). Jos taas uusi yhdyssana muodostetaan suoraan kahdesta sanasta yhdistämällä ilman sanaliittomaista välivaihetta, sana termistyy helpommin (*helppokaava*), jolloin myös sen alkuosa kiteytyy todennäköisemmin tiettyyn muotoon. Toisesta näkökulmasta tarkasteltuna voidaan sanoa, että juuri alkuosan taipumattomuus tekee sanasta termimäisemmän (R. Lehtinen 1984: 518; Eronen 1996: 12; T. Kolehmainen 2014: 159).

Kieliopillis-kielenhuollollisesta näkökulmasta osiltaan kongruoiviksi tulkittavien yhdyssanojen määrä (tutkimusaineistossa 6/28019 kpl eli n. 0,02 %) on uudessa yhdyssanastossa häviävän pieni. Kuitenkin tässä pienessä joukossa olevat täysin uudet muodosteet osoittavat, että osiltaan kongruoivien uusien yhdyssanojen synty suomen kieleen on edelleen mahdollista. Sanaliitoista tiivistymisen lisäksi myös erilaiset analogiat vanhoihin muodosteisiin voivat tukea kongruenssin säilymistä.

Kaikkien tutkimusaineistossa olevien osiltaan valinnaisesti kongruoivien yhdyssanojen kongruenssin voidaan osoittaa motivoituneen joko sanaliitosta tiivistymisen, yksittäiseen sanaan perustuvan analogian (vrt. *mustasuola* ja *mustapippuri*) tai molempien kautta. Produktiivisuuden ja esimerkiksi analogian kautta syntyneiden yksittäistapausten välille ei voida vetää tarkkaa rajaa, mutta mikäli jokaiselle tiettyä rakennetta noudattavalle sanalle voidaan osoittaa jokin muu motivaatio kuin produktiivisen sananmuodostusmallin seuraaminen, on rakenteen produktiivisuus vähintään kyseenalaistettava (ks. myös Rintala 1978: 153, 162; Bauer 2001: 4–5). Näyttää siis siltä, että osiltaan kongruoivien yhdyssanojen rakennetta ei voida ainaakaan enää pitää suomen kielessä produktiivisena.

ISK:n (2004: § 420) mukaan PS:ssa⁷⁸ on 93 osiltaan kongruoivaa adjektiivialkuista yhdyssubstantiivia (näistä aina kongruoiviksi on mainittu 47 kpl ja valinnaisesti kongruoiviksi 46 kpl). Vastaavia osiltaan kongruoimattomia tapauksia on 494 kpl. CD-PS:ssa on kaikkiaan 51 356 kahdesta leksikaalisesta elementistä koostuvaa yhdyssubstantiivia (Jussi Niemi 2009: 250–251). Osiltaan kongruoivien osuus tästä joukosta on 0,18 %. NS:ssa on ISK:n (mp.) mukaan 276 osiltaan kongruoivaa yhdyssanaa. NS:n sisältämien kahdesta leksikaalisesta elementistä koostuvien yhdyssubstantiivien määrästä ei ole tietoa, mutta NS:n kaikkiaan noin 130 000 yhdyshakusanasta (Saukkonen 1973: 338; tosin Saukkonen ei tuo esiin analyysissaan käyttämiään yhdyssanan kriteerejä) osiltaan kongruoivien tapausten osuus on 0,21 %. Pelkistä 2-osaisista yhdyssubstantiiveista osuus on tietysti vielä suurempi. Sanakirjojen valossa osiltaan kongruoivien yhdyssanojen määrä näyttää siis vähenneen suomen kielessä NS:n ja CD-PS:n koostamisen välillä.

Omassa tutkimusaineistossani osiltaan kongruoivien yhdyssanojen osuus on selvästi pienempi kuin kummassakaan esitellyssä sanakirjassa. Kaikista aineiston yhdyssanoista osiltaan kongruoivien osuus on 0,02 % ja 2-osaisista yhdyssubstantiiveista 0,03 %. Vertailun helpottamiseksi tiedot osiltaan kongruoivien yhdyssanojen osuudesta NS:ssa, PS:ssa ja uudessa yhdyssanastossa on koottu taulukkoon 14.

78 PS:n ja CD-PS:n sisältö on pääosin sama (M. Lehtinen 1997).

TAULUKKO 14. Osiltaan kongruoivat yhdyssanat tutkimusaineistossa sekä NS:ssa ja PS:ssa

	Osiltaan kongruoivien osuus kaikista yhdyssanoista	Osiltaan kongruoivien osuus 2 leksikaalista osaa sisältävistä yhdyssubstantiiveista
NS (<i>n</i> = 130 000)	0,21 %	?
(CD-)PS (<i>n</i> = 51 356)	?	0,18 %
TT (<i>n</i> = 28 019/20 754)	0,02 %	0,03 %

Koska osiltaan kongruoivien yhdyssanojen määrä on hyvin pieni sekä uudessa yhdyssanastossa että sanakirja-aineistoissa, pidemmälle vietyjen tilastollisten tarkastelujen tekeminen ei ole järkevää. Sitä, syntykö osiltaan kongruoivia sanoja vähemmän kuin ennen, on mahdotonta arvioida tämän vertailun perusteella, sillä ei ole tietoa, kuinka pitkän ajan kuluessa sanakirjoihin otetut osiltaan kongruoivat yhdyssanat ovat syntyneet tai kuinka paljon sanakirjaan koskaan päätyttömiä osiltaan kongruoivia yhdyssanoja kielessä on milloinkin ollut käytössä. Tämä tutkimus kuitenkin vahvistaa aiemmissä tutkimuksissa tehtyä havaintoa siitä, että uudet yhdyssanat asettuvat vain harvoin osiltaan kongruoivien yhdyssanojen luokkaan. Koska osien kongruenssi on jo valmiiksi harvinainen piirre suomen kielen yhdyssanoissa, sen voi olettaa edelleen harvinaistuvan (ks. myös Bauer 2001: 2–4; T. Kolehmainen 2014: 157–165). Aiempi tutkimus on osoittanut, että kongruenssi on katoamassa myös monista vanhastaan kongruoiveista yhdyssanoista (esim. *isosisko* : *isonsiskon* ~ *isosiskon*; R. Lehtinen 1984; Eronen 1996).

On mahdollista, että kieliopin toistaiseksi suoma mahdollisuus osiltaan kongruoivien yhdyssanojen luomiseen ja käyttöön katoaa suomen kielestä ajan kuluessa kokonaan. Piirteen säilymistä elinvoimaisena tukee kuitenkin se, että monet yksittäiset osiltaan kongruoivat yhdyssanat tai niiden tietyt muodot ovat varsin frekventtejä. Monien yhdyssanojen osien kongruenssi vaihtelee sen mukaan, mihin sijaan sanaa taivutetaan. Yleisintä osien kongruenssi on kieliopillisissa sijoissa (esim. *omantunnon*). Esimerkiksi vakiintuneiden sanontojen seurauksena kongruenssia voi esiintyä sanakohtaisesti paljon myös muissa sijoissa (esim. *omallatunnolla*). Mikäli *omantunnon* kaltaiset frekventit muodosteet joskus korvautuvat toisilla tai muuttuvat osiltaan kongruoimattomiksi, ei uusille osiltaan kongruoiville yhdyssanoille löydy enää edes analogista mallia.

Yksi mahdollinen syy osiltaan kongruoivien yhdyssanojen katoon on kirjallisen viestinnän lisääntyminen. Yhdyssanojen oikeinkirjoitus tuottaa monille vaikeuksia, ja osiltaan kongruoivat yhdyssanat voivat vaikeuttaa yhdyssanojen oikeinkirjoitussääntöjen tulkitsemista entisestään. Kun yhdyssanan alkuosa jätetään taivuttamatta, normatiivisen yhdyssanan ja sanaliiton välille saadaan selkeämpi ero. Myös kielenhuolto sallii uudenlaisen taivutuksen, sillä kieli on tässä asiassa selvästi muuttunut (R. Lehtinen 1984; Eronen 1996). Kaiken kaikkiaan osiltaan valinnaisesti kongruoivien yhdyssanojen kongruoimattomuus alkaa Erosen (1996: 12) mukaan olla nykysuomalaisten kielitajussa yleisempää kuin kongruenssi. Nykytilanteesta ei ole

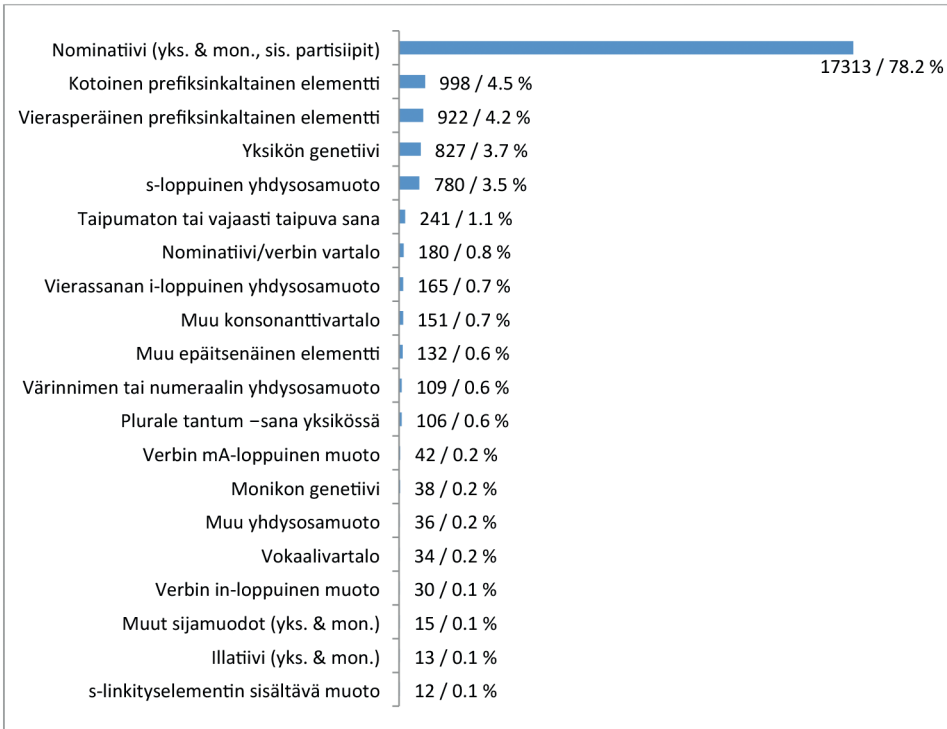
olemassa tutkittua tietoa, eikä asiaa ole toistaiseksi tutkittu esimerkiksi todellista kielenkäyttöä heijastavien korpusten avulla.

Toinen mahdollinen syy on yleiskielen sanaston muuttuminen termimäisemmäksi. Esimerkiksi murteissa yhdyssanat ovat useimmiten sanaliitoista kiteytyneitä, kun taas yleiskieleen – jossa on kaiken kaikkiaan enemmän yhdyssanoja kuin murteissa – kuuluu paljon selitteen- tai määritelmänluonteisia alakäsitteiden nimiä (Tuomi 1989: 28). Termimäisten yhdyssanojen puolestaan on todettu olevan harvoin osiltaan kongruoivia (R. Lehtinen 1984; Eronen 1996: 12). Nimenomaan yleiskielessä osiltaan kongruoivien yhdyssanojen väheneminen voikin olla seurausta kielimuodon muuttumisesta enemmän erikoiskielisen viestinnän suuntaan.

Kolmas syy osiltaan kongruoivien yhdyssanojen vähenemiselle voi olla kongruenssin redundanssi. Yhdyssanan molempien osien taivuttaminen ei tuo ilmaukseen lisäinformaatiota, minkä voisi pienimmän vaivan periaatteen mukaisesti olettaa ohjaavan kieltä kohti kongruenssista luopumista. Voi olla, että jotkut kokevat toiston myös esteettisesti epämiellyttäväksi. Redundanssi ei kuitenkaan yksioikoisesti aiheuta kielen muutosta. Kieliopillisesti redundantilta vaikuttavalla piirteellä voi myös olla esimerkiksi viestin välittämisen varmistamiseen liittyviä tehtäviä (Keller-Nerlich 1994: 104–106). Vaikka kongruenssi yleensä puuttuu yhdyssanoista eikä piirre vaikuta ainakaan välttämättömältä sen enempää kieliopillisten kuin viestin välittämiseen tai esimerkiksi tyyliarvoon liittyvien syiden vuoksi, kongruenssi voi kuitenkin vaikuttaa yhdyssanojen prosessointiin. Kirjoitettua suomen kieltä tutkittaessa on todettu, että kongruoivan adjektiivin sisältämien substantiivilausekkeiden (*suuressa talossa*) prosessointi on sujuvampaa kuin kongruoimattoman adjektiivin sisältävien (*pikku talossa*) (Vainio-Hyönä-Pajunen 2003, 2008; Hyönä-Vainio 2009: 725–726). Koska myös yhteen ja erikseen kirjoittaminen vaikuttaa prosessointiin (ks. esim. Inhoff-Radach-Heller 2000) eikä erityisesti yhdyssanan osien kongruenssin vaikutuksesta sanamuodon tulkinnan sujuvuuteen ole tutkittua tietoa, ei voida sanoa varmasti, onko yhdyssanan osien kongruenssi todella redundanttia vai ei.

5.3.4 Yhteenveto

Kuviossa 4 on esitetty yhteenveto tutkimuksen alkuosan muotojakaumaa koskevista tuloksista. Kuvioon on yhdistelty tietoja kaikista luvun 5.3.2 alaluvuista. Kaikki pienimmät alakategoriat eivät kuitenkaan ole näkyvissä kuviossa.



KUVIO 4. Yhteenveto tutkimusaineiston alkuosan hienojakoisesta muotojakaumasta ($n = 22\,144$)

Uudessa yhdyssanastossa ylivoimaisesti yleisin alkuosan muoto on yksikön nominatiivi. Nominatiivin jälkeen yleisimmät yhdyssanan alkuosan muodot ovat kotoiset ja vierasperäiset prefiksinkaltaiset yhdysosat, jotka molemmat ovat vanhastaan tärkeänä pidettyä genetiivialkua käytetympiä. Tämä tutkimus osoittaa, että prefiksinkaltaiset yhdysosat on viimeistään nyt syytä ottaa huomioon tärkeänä suomen kielen sananmuodostuskeinona. Jatkossa on syytä huomioida myös aiemmin huomiotta jäänyt *-inen*-loppuisten vierassanojen *-i*-loppuinen yhdysosamuoto. Tyyppi on melko harvinainen, mutta sen muodostus on säännönmukaista.

Genetiivialkuisuus on uusien yhdysadjektiivien joukossa harvinaisempaa kuin sanakirja-aineistossa (vastaavasti nominatiivialkuisuus on yleisempää; yhdysadjektiivit ovat ainoa ryhmä, josta on saatavissa luotettavaa aiempaa tietoa), ja nominatiivia käytetään uudessa yhdyssanastossa myös sellaisissa sanoissa, joissa genetiivi tuntuisi luontevammalta. Todennäköinen syy genetiivialkujen harvinaistumiseen on kirjallisen viestinnän lisääntyminen yhdistettynä kielenkäyttäjien epävarmuuteen oikeinkirjoitussäännöistä. Nominatiivialulla on myös helpompaa osoittaa ilmauksen käsitteellinen yhteenkuuluvuus.

Suomen kielen yhdyssanoissa käytössä olevia mutta uudessa sanastossa marginaaliin jääviä alkuosan muotoja ovat nominin monikkovartaloa vastaava *-o*-loppuinen muoto ja *s*-linkityselementin sisältävä muoto. Näitä muotoja käytetään uudessa

sanastossa vain sellaisista alkuosista, joiden yhteydessä ne ovat vanhastaankin esiintyneet. Myös muita sijamuotoja kuin nominatiivia ja genetiiviä edustavat alkuosat ovat hyvin harvinaisia vaikka sananmuodostuksessa edelleen täysin mahdollisia.

Suomen kieleen syntyy hyvin vähän uusia osiltaan kongruoivia yhdyssanoja. Uusien osiltaan kongruoivien yhdyssanojen synty on kuitenkin mahdollista sanaliittojen tiivistymisen ja analogian kautta. Tyypin marginaalisuuden ja aiemman tutkimuksen vähäisyyden vuoksi ei voida varmasti sanoa, onko osiltaan kongruoivien yhdyssanojen tyyppi harvinaistumassa suomen kielessä. Todennäköisiä syitä tyypin harvinaisuudelle uudessa sanastossa ovat osiltaan kongruoivien yhdyssanojen harvinaisuus vakiintuneessakin sanastossa, kirjallisen viestinnän lisääntyminen, sanaston termistyminen sekä yhdyssanan osien kongruenssin redundanssi.

6 Rakennepiirteet ja yhdyssanan muodostaminen

Luvun 5 kvalitatiivinen analyysi nosti selvästi produktiivisten rakenteiden lisäksi esiin sellaisia rakenteita, jotka vaikuttavat olevan produktiivisempia johtamisen kuin yhdistämisen kautta (esim. yhdysverbit, ks. luku 5.2.2), sekä rakenteita, joiden produktiivisuus näyttää olevan seurausta enemmän yksittäisen rakenteessa käytettävän sanan kuin itse rakenteen suosiosta (esim. pronomiaalikeset yhdyssanat, ks. luku 5.2.4). Tässä luvussa käsitellän tarkemmin eri rakenteita noudattavien yhdyssanojen muodostustapaa (luku 6.1) ja leksikaalista monipuolisuutta (luku 6.2). Yhdyssanan muodostustavan tarkastelu poikkeaa tutkimuksen muuten synkronisesta linjasta.

6.1 Muodostustapa

Uusia yhdyssanoja muodostetaan sekä yhdistämällä että johtamalla (Koivisto [2013: 263–264] nimittää johtamalla muodostettuja yhdyssanoja *johdosyhdyssanoiksi*). Koska tämä tutkimus on luonteeltaan synkroninen, tutkimusaineistoni on koostettu pelkästään sanan muotoa kriteerinä käyttäen ja se sisältää sekä yhdistämällä että johtamalla muodostettuja yhdyssanoja.

Uusien yhdyssanojen aineksina voidaan käyttää perussanoja, johdoksia tai aiemmin muodostettuja yhdyssanoja. Periaatteessa muodostustapoja voidaan erottaa neljä: 1) suoraan yhdistämällä muodostetut (esim. *ihminen + banaani > ihmisbanaani*), 2) aiemmasta yhdyssanasta johdetut (*nuorisokulttuuri + -mainen > nuorisokulttuurimainen*), 3) sanaliitosta tiivistyneet (*maalle muutto > maallemuutto*) ja 4) sanaliitosta johdetut (*oikea silmä > oikeasilmäinen*) (ks. myös 2.2.3; tarkempaa jaottelua käyttää esim. Levonen 2006: 5).⁷⁹ Yhdyssanojen muodostuksessa samaan lopputulokseen voidaan päätyä useaa eri reittiä (ks. myös luku 4.1). Kun sanan käyttöhistoriaa ei ole tarkoin dokumentoitu, rajanveto eri muodostustapojen välillä voi

⁷⁹ Syvärakenteita olettavassa generativistisessä yhdyssanatutkimuksessa erotetaan yleensä omaksi ryhmäkseen deverbaalit yhdyssanat (esim. *riistää piste > pisteriisto*; ks. tarkemmin esim. Mäkisalo 2000a: 15–16). Koska tässä tutkimuksessa yhdyssanojen taustalle ei oleteta syvärakenteita, esimerkiksi sana *pisteriisto* on luokiteltu kahdesta sanasta (*piste + riisto*) suoraan yhdistetyksi yhdyssanaksi.

olla vaikeaa. Esimerkiksi yhdyssana *palstanvaltaaja* voidaan tulkita joko kahdesta sanasta suoraan yhdistetyksi (*palsta + valtaaja > palstanvaltaaja*, alkuosan genetii-vimuoto osoittamassa osien suhdetta) tai sanaliiton osista tiivistymällä yhdistetyksi (*palstan valtaaja > palstanvaltaaja*).

Käytän tässä luvussa kaksivaiheista analyysia. Analyysin ensimmäinen vaihe on tilastollinen, ja erotan siinä toisistaan vain kaksi muodostuskategoriaa: yhdistämisen ja johtamisen. Kahden kategorian käyttö on tilastollisessa analyysissa selvempää, sillä edellisessä kappaleessa esitellyt neljä kategoriaa limittyvät keskenään ja tarkkojen rajanvetojen teko on siksi ongelmallista. Luokittelussa yhdistämisen alle kuuluvat suora yhdistäminen ja sanaliitosta tiivistyminen. Lisäksi olen laskenut yhdistämisen alle sanaliitosta johtamisen, jossa yhdistyvät sekä yhdistäminen että johtaminen. Johdettujen ryhmään jäävät näin ollen ainoastaan aiemmista yhdyssanoista johdetut muodosteet. Tällaisen dikotomisen analyysin avulla voidaan erottaa ainoastaan johtamisen kautta syntyneet yhdyssanat niistä, joiden muodostusprosessissa on hyödynnetty (edes yhtenä keinona) yhdistämistä.

Tilastollisessa analyysissa olen käyttänyt luokittelun tukena ensisijaisesti tietoa sanan merkityksestä. Esimerkiksi sana *kiintopisteetön* viittaa sellaiseen, jolla ei ole kiintopistettä (ei sellaiseen, jolla ei ole pistettä mutta joka on kiinto/kiinteä; ts. määriteosa ei enää määritä edusosaa suoraan). Näin ollen muodoste on tulkittava johdokseksi aiemmasta *kiintopiste*-yhdyssanasta. Tarvittaessa olen huomioinut luokittelussa myös tiedon sanan ja sen taustalla mahdollisesti olevien kantasanojen historiasta, muodostusmalleista ja käyttöfrekvensseistä (lähteinä mm. NSST, eri-ikäiset sanakirjat, oma kielitieto ja internet). Historiaan ja muodostusmalliin nojaten olen luokitellut yhdistämällä muodostetuksi esimerkiksi sanan *kuivapelata* (esim. 29). Sana on ilmeisesti muodostettu aiemman *kuivanarkkaroida*-yhdyssanan mallin mukaan, eikä se näin ollen todennäköisesti ole johdos esimerkiksi sanasta *kuivapeli*.

- (29) Pelaajat tuntevat toisensa sen verran hyvin, että jos sana lähtee kiertämään, että tukiryhmässä *kuivapelataan*, kukaan ei tule enää seuraavalla kerralla. Kuivapelaaminen tarkoittaa pelaamiseen liittyvää tunnelmointia. Huumeiden väärinkäyttäjillä ilmaus on kuivanarkkarointi eli huume-filistely. Weidner on itse aikaisemmin työskennellyt päihderiippuvaisten kanssa vankiloissa. [TT <- HS 29.12.2008, Juhana Pettersson, s. C 8]

Monet aiemmista yhdyssanoista johdetuiksi luokittelemani sanat ovat sellaisia, että ne voidaan muodostaa eri käyttökerroilla yhä uudestaan (esim. *kotivärjätä*; ks. myös luku 4.1). Jos sanalle on kuitenkin osoitettavissa aiempi yhdyssanamuotoinen kantasana (esim. *kotivärjäys*) eikä muodosteiden yhdistämiselle ole mitään estettä (esim. erilainen merkitys), olen luokitellut nuoremman muodosteen johtamalla syntyneeksi.

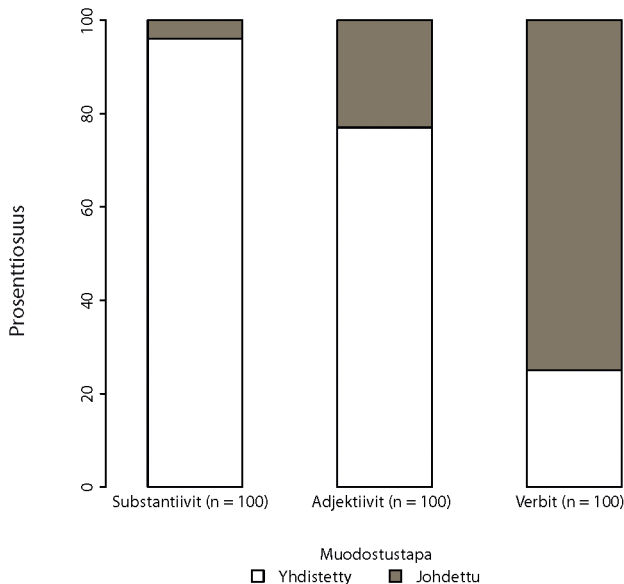
Analyysin toinen vaihe on luonteeltaan kvalitatiivinen, ja siinä tarkastelen lähemmin reittejä, joita pitkin yksittäisiin muodosteisiin on päädytty. Kvalitatiivisen analyysin avulla saadaan tietoa siitä, mitä rakenteita noudattavat sanat muodostetaan tyypillisesti suoraan kahdesta sanasta yhdistämällä ja mitkä puolestaan sanaliitosta johtamisen tai tiivistymisen kautta.

Luvussa 5.2.2 totesin, että eri sanaluokkia edustavista yhdyssanoista johtaminen näyttää olevan tyypillinen muodostuskeino varsinkin yhdysverbeille (ks. myös Rahtu 1984; Levonen 2006). Luvussa 5.3.2.6 puolestaan totean, että sanaliitosta tiivistyminen näyttää olevan tyypillistä nimenomaan muulla sijalla kuin nominatiivilla tai genetiivillä alkaville yhdyssanoille. Tässä luvussa käsittelen tarkemmin sekä sanaluokan (6.1.1) että alkuosan muodon (6.1.2) yhteyttä yhdyssanan muodostustapaan.

Koska yhdyssanojen muodostustavan tarkastelu vaatii tuekseen erityisen tarkkaa kvalitatiivista analyysia, tässä luvussa ei ole mahdollista käsitellä koko tämän tutkimuksen laajaa aineistoa. Siksi käsittelen uusien yhdyssanojen muodostustapaa kahden pienemmän osa-aineiston kautta. Osa-aineistot on koostettu sanaluokkaa ja alkuosan muotoa perusteena käyttäen satunnaisotannalla osa-aineistosta 3, joka sisältää vain kaksiosaisia yhdyssanoja ($n = 22\ 144$; osa-aineistot on esitelty tarkemmin luvussa 4.3). Muodostustavan ja sanaluokan yhteyttä käsittelevässä luvussa (6.1.2) käytetyssä otoksessa on 300 yhdyssanaa, joista 100 on satunnaisesti valittuja substantiiveja, 100 adjektiiveja ja 100 verbejä. Muodostustavan ja alkuosan muodon yhteyttä käsittelevässä luvussa (6.1.2) käytetyssä otoksessa on 228 yhdyssanaa, joista 100 satunnaisesti valittuja nominatiivialkuisia ja 100 genetiivialkuisia sanoja. Lisäksi aineistossa ovat kaikki aineiston 28 muussa sijamuodossa olevan alkuosan sisältävät sanat (ks. luku 5.3.2.6).

6.1.1 Muodostustapa ja sanaluokka

Uusien yhdyssanojen muodostustavan ja sanaluokan suhdetta on havainnollistettu kuviossa 5.



KUVIO 5. Tutkimusaineiston yhdyssanojen muodostustavan ja sanaluokan suhde ($n = 300$)

Kuten kuviosta 5 nähdään, substantiiveista valtaosa (96 %) on yhdistämällä muodostettuja. Adjektiiveissa yhdistettyjen osuus on selvästi pienempi (77 %). Verbeistä yhdistämällä muodostettuja on vain 25 %. Jakauma osoittaa selvästi todeksi sen, että yhdysverbit ovat muita yhdyssanoja useammin aiemmista yhdyssanoista johdettuja (ks. myös luku 5.2.2 ja esim. Vesikansa 1989: 256). Kuitenkin yhdysverbejäkin muodostetaan myös yhdistämällä.

Sanaluokkien väliset erot ovat tilastollisesti merkitseviä (Pearsonin khin neliö riippumattomuustestin $\chi^2 = 120$, $n = 300$, $df = 2$, $p < 0,001^{***}$). Sovitettujen standardoitujen jäännösten avulla tehdyn jatkotarkastelun perusteella eniten kaikkien ryhmien yhteisestä keskiarvosta poikkeavat yhdysverbit, joiden joukossa on tilastollisesti merkitsevästi enemmän johtamalla ja vähemmän yhdistämällä muodostettuja sanoja (jäännökset = -5 ja 7). Substantiivien ryhmässä puolestaan on tilastollisesti merkitsevästi enemmän yhdistämällä muodostettuja ja vähemmän johtamalla muodostettuja tapauksia kuin uusissa yhdyssanoissa keskimäärin (jäännökset = 3,7 ja -5,1). Adjektiivit eivät poikkea tilastollisesti merkitsevästi kaikkien ryhmien yhteisestä keskiarvosta (jäännökset 1,4 ja -1,9).

Pareittain tehdyn jatkotarkastelun mukaan kaikki kolme ryhmää eroavat tilastollisesti merkitsevästi toisistaan (substantiiveille ja adjektiiveille tehdyn khin neliö -testin $\chi^2 = 15,46$, $n = 200$, $df = 1$, Bonferronin–Holmin korjauksella tarkistettu $p < 0,001^{***}$; adjektiiveille ja verbeille tehdyn khin neliö -testin $\chi^2 = 54,1$, $n = 200$, $df = 1$, tarkistettu $p < 0,001^{***}$; substantiiveille ja verbeille tehdyn khin neliö -testin $\chi^2 = 105,5$, $n = 200$, $df = 1$, tarkistettu $p < 0,001^{***}$).

Suurin osa suomen kielen uusista yhdyssanoista on substantiiveja (ks. luku 5.2.2). Koska yhdyssubstantiiveista jopa 96 % on yhdistämällä muodostettuja, suurimman osan suomen kielen uusista yhdyssanoista ylipäänsä voidaan sanoa syntyvän yhdistämällä. Tutkimusotoksen yhdistämällä muodostetuista yhdyssubstantiiveista lähes kaikki on muodostettu liittämällä kaksi sanaa suoraan yhteen (esim. *pelaajakäytävä* 'käytävä, jota pitkin pelaajat kulkevat pukuhuoneesta kentälle', *kännykkälupa* 'kännykän kautta hankittu kalastuslupa'). Tämä on odotuksenmukaista, sillä suurimassa osassa uusista yhdyssanoista alkuosa on nominatiivimuotoinen substantiivi (ks. luvut 5.2.4 ja 5.3.1), ja suomen kielessä kaksi substantiiviva voi yleensä muodostaa sanaliiton vain, jos niistä ensimmäinen on jossakin muussa sijassa kuin nominatiivissa. Niinpä tutkimusotoksen substantiiveista vain jotkin genetiivialkuiset muodosteet voidaan tulkita joko suoraan yhdistämällä tai sanaliitosta tiivistymällä muodostetuiksi (esim. *arvonluonti*, *kodinvaihto*; muulla sijalla alkavia ei sattunut tähän pieneen satunnaisesti valittuun osa-aineistoon; käsittelen alkuosan muodon ja muodostustavan yhteyttä seuraavassa luvussa 6.1.2). Johtamalla syntyneet uudet yhdyssubstantiivit on tyypillisesti johdettu aiemmista yhdysubstantiiveista (esim. *ympäristökasvatus* > *ympäristökasvattaja*, myös ns. takaperoinen johtaminen on mahdollista, esim. *hohtokeilaus* > *hohtokeila*).

Aiemmista yhdyssubstantiiveista johtaminen on tyypillistä myös muille sanaluokille. Johtamalla muodostetuista adjektiiveista (23 % kaikista adjektiiveista) kaikkien taustalla on yhtä lukuun ottamatta nimenomaan yhdyssubstantiivi (esim.

nuorisokulttuuri > *nuorisokulttuurimainen*, *tunneäly* > *tunneälykäs*; poikkeus adverbistä *suutuspäissään* johdettu *suutuspäinen*). Yhdistämällä muodostetuille uusille yhdysadjektiiveille on yhdyssubstantiivien tapaan tyypillistä suoraan kahdesta sanasta yhdistäminen (*ei-aristokraattinen*, *perusonnellinen*). Yhdysadjektiiveille on tavallista kuitenkin myös sanaliitoista tiivistyminen sekä yhdistämisen ja johtamisen limittyminen. Varsinkin *-inen*-loppuiset adjektiivit ovat usein sekä sanaliitoista tiivistymisen että johtamisen kautta syntyneitä (esim. *löysäkielinen*, *pienikustannuksinen*). Selvimmin yhdistäminen ja johtaminen limittyvät sellaisissa muodosteissa, joiden jälkiosa on epäitsenäinen *-inen*-adjektiivi eikä taustalla ole selvää sanaliittoa (esim. *pikakiristeinen* ja *poistogeeninen* 'sellainen, josta on poistettu DNA:ta').

Uusista yhdysverbeistä suurin osa on johtamalla muodostettuja (75 % kaikista verbeistä). Johdettujen yhdyssubstantiivien ja *-adjektiivien* tapaan myös johdetuista yhdysverbeistä suurin osa on johdettu nimenomaan yhdyssubstantiiveista (esim. *hintalappu* > *hintalaputtaa*, *vapaamatkustaja* > *vapaamatkustaa*). ISK:n (§ 430) mukaan johdetun yhdysverbin kanta on useimmiten *-inen*-loppuinen adjektiivi. Uusien yhdyssanojen kohdalla tämä ei näytä ainakaan enää pitävän paikkaansa, sillä tutkimusotoksessa on vain viisi *-inen*-adjektiiveista johdettua yhdysverbiä (esim. *kansanomainen* > *kansanomaistaa*, *vieraanvarainen* > *vieraanvaraistaa*). Substantiivien ja *-inen*-adjektiivien lisäksi yhdysverbejä johdetaan myös muista adjektiiveista sekä aiemmista yhdysverbeistä (*upporikas* > *upporikastua*, *ruiskumaalata* > *ruiskumaalauttaa*, *salakuljettaa* > *salakuljettua*).

Yhdysverbejä on vanhastaan pidetty vieraina suomen kielen rakenteelle ja niiden asemasta on käyty vuosien saatossa paljon keskustelua kielenhuollossa (ks. esim. Sadeniemi 1955: 300; Saarimaa 1971 [1947]: 88–90; Leskinen 1983: 571; Rahtu 1984; Grönros 1994). Suomen kielen rakenne ei kuitenkaan sinänsä estä yhdysverbien muodostamista, vaan rajoitukset ovat lähinnä tyylillisiä ja kielenhuollon asettamia. 1900-luvulla kielenhuoltajat ovat neuvoneet välttämään varsinkin takaperoisesti johdettuja yhdysverbejä (esim. *vakuuttaa* > *vakuutus* mutta *palovakuutus* > *palovakuuttaa*; esim. Saarimaa 1971 [1947]: 88–89; tyypistä ks. myös Rahtu 1984), mutta niiden tarpeellisuutta on myös puolustettu (esim. Vesikansa 1989: 257–258; ks. myös Rahtu 1984: 422–423, 426). Nykyään yhdysverbeihin suhtaudutaan huomattavasti aiempaa sallivammin, ja päätökset yksittäisten muodosteiden hyväksyttävyydestä perustetaan ennemmin sanojen vakiintuneisuuteen ja tarpeellisuuteen kuin johtosuuntaan (ks. esim. Grönros 1994). Uudessa yhdyssanastossa yleistä on sekä suoraan että takaperoisesti johtaminen (suoraan johdettuja esim. *korppujauho* > *korppujauhottaa* ja *virkamies* > *virkamiehistyä*; takaperoisesti johdettuja esim. *etäopetus* > *etäopettaa* ja *tulosvaroitus* > *tulosvaroittaa* sekä em. *vapaamatkustaja* > *vapaamatkustaa*).

Yhdysverbien on myös monissa aiemmissä tutkimuksissa todettu olevan usein toisista yhdyssanoista johdettuja (Vesikansa 1989: 256–258; Tuomi 1995: 154; Levonen 2006). Tämän tutkimuksen aineistossa johtamalla muodostettujen yhdysverbien osuus (75 %) on kuitenkin selvästi suurempi kuin aiemmissä kvantitatiivisissa

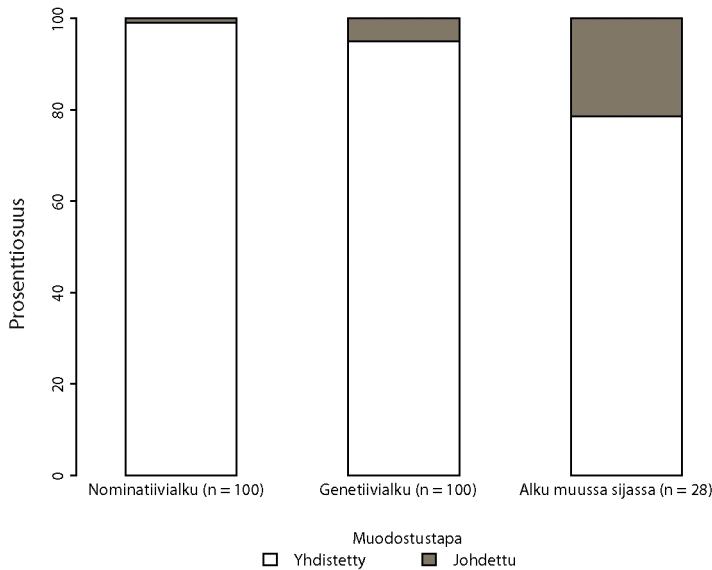
tutkimuksissa. HS 2000–2001 -korpuksen verbaaliyhdyssanoja tutkineen Levosen (2006: 57, 114–122) mukaan verbaaliyhdyssanoista on muodostunut yhdistymällä 55,4 % ja johtamalla 44,6 %. Erot verbiparadigman eri muodoissa esiintyvien sanojen sekä eri-ikäisten sanojen välillä ovat kuitenkin suuret. Levosen aineistossa aktiivin finiittimuodoissa esiintyvistä yhdysverbeistä 68,5 % (esim. *kerskakuluttaa*, *penkkiurheilla*) ja passiiviin finiittimuodoissa esiintyvistä 57,7 % (esim. *kielikylvetään*, *lisäverotetaan*) on johtamalla muodostettuja. Infinitiivimuodoista johtamalla on muodostettu 55 % (esim. *potkukelkkailemalla*, *uudelleenjärjestellä*). Vähiten johtamalla muodostettuja on partisiippimuodoissa (37,7 %, esim. *allekirjoitettu*, *radio-ohjattava*; omassa tutkimuksessani partisiipit on erotettu omaksi kategoriakseen eikä niitä ole laskettu verbien ryhmään lainkaan). Levosen aineisto sisältää myös vanhaa sanastoa⁸⁰, joten ainakin näiden kahden tutkimuksen vertailu viittaa siihen, että yhdysverbien muodostaminen johtamalla on nykyään jopa yleisempää kuin ennen.

Tutkimusotoksen yhdistämällä muodostetuille yhdysverbeille on tyypillistä osien suora yhdistäminen (*ilmasoittaa*, *pintaviimeistellä*). Suoraan yhdistetyn yhdysverbin ensimmäinen osa on usein prefiksinkaltainen (*etälukea*, *salanaida*). Vanhastaan tyypillisenä yhdysverbien syntytapana pidetty sanaliitoista tiivistyminen on sen sijaan uudessa sanastossa harvinaista. Levonen (2006: 35) esittää tyyppin harvinaisuuden syyksi sitä, että finiittimuotoisen verbin määritteet eivät enää nyky-suomessa ole verbien edellä (ennen myös verbin määritteet ovat olleet pääsanansa edellä, ks. Häkkinen 1987: 20–21). Sanaliiton tiivistyminen yhdysverbiksi edellyttää siis nyky-suomessa joko sanajärjestyksen kääntämistä tai infiniittistä välivaihetta (esim. *rakentaa alas* > *alasarakentaa* 'purkaa niin, että materiaaleja voidaan käyttää uudelleen' ja *hitsautua yhteen* > *yhteenhitsautua*).

6.1.2 Muodostustapa ja alkuosan muoto

Uusien yhdyssanojen muodostustavan ja alkuosan muodon suhdetta on havainnollistettu kuviossa 6.

80 Levosen aineiston yksifrekvenssisistä verbaaliyhdyssanoista (voidaan parhaiten rinnastaa uusiin yhdysverbeihin, ks. Baayen 1989, 1992, 1993; Baayen–Lieber 1992; ks. myös luku 3.3.2) johtamalla muodostettuja on 41,4 %. Tässä osuudessa ovat kuitenkin mukana myös partisiippimuodot, joita on n. puolet aineistosta. (Levonen 2006: 132.)



KUVIO 6. Tutkimusaineiston yhdyssanojen muodostustavan ja alkuosan muodon suhde ($n = 228$)

Kuten kuviosta 6 nähdään, suurin osa uusista yhdyssanoista on yhdistämällä muodostettuja riippumatta niiden alkuosan muodosta⁸¹. Eniten yhdistämällä muodostettuja on nominatiivialkuisten sanojen ryhmässä (99 %). Seuraavina tulevat genetiivialkuiset, joista yhdistämällä on muodostettu 95 %. Muussa sijamuodossa olevalla alkuosalla alkavista sanoista 79 % on yhdistämällä muodostettuja. Kaiken kaikkiaan tässä luvussa käytetyssä otoksessa on vain 12 johtamalla muodostettua yhdyssanaa. Tämä vahvistaa edellisen luvun (6.1.1) havaintoa siitä, että suurin osa suomen kielen uudista yhdyssanoista muodostetaan yhdistämällä (tosin yhdysverbit ovat poikkeus tähän sääntöön).

Vaikka yhdistäminen on yhdyssanan yleisin muodostustapa alkuosan muodosta riippumatta, ryhmien välillä on kuitenkin selkeitä eroja, ja erot ovat myös tilastollisesti merkitseviä (Fisherin tarkan testin $n = 228$, $p < 0,001^{***}$). Sovitettujen standardoitujen jäännösten avulla tehdyn jatkotarkastelun mukaan muulla sijalla alkavien ryhmässä on tilastollisesti merkitsevästi enemmän johtamalla muodostettuja sanoja kuin uusissa yhdyssanoissa keskimäärin (jäännös = 3,7). Muut erot eivät ole jäännösten perusteella tilastollisesti merkitseviä.

Parittaisen jatkotarkastelun mukaan nominatiivi- ja genetiivialkuisten sanojen muodostustavassa ei ole tilastollisesti merkitsevää eroa (Fisherin tarkan testin

81 Tässä luvussa käytetyssä satunnaisotoksessa on 210 substantiivia (n. 92,1 % otoksesta). Tämä vastaa melko hyvin koko tutkimusaineiston jakaumaa (substantiivien osuus 92,8 %, ks. luku 5.2.2). Kuten edellisessä luvussa (6.1.1) totesin, substantiiveista suurin osa on yhdistämällä muodostettuja. Tämän luvun tutkimusotoksessa olevista 12:sta johtamalla muodostetusta sanasta kuitenkin 10 on substantiiveja ja vain 2 verbejä.

Bonferronin–Holmin korjauksella tarkistettu $p = 0,21$), mutta muiden ryhmien väliset erot ovat tilastollisesti merkitseviä (genetiivillä ja muulla sijamuodolla alkaville tehdyn Fisherin tarkan testin tarkistettu $p = 0,028^*$; nominatiivilla ja muulla sijamuodolla alkaville tehdyn Fisherin tarkan testin tarkistettu $p = 0,003^{**}$).

Nominatiivialkuisille yhdyssanoille on tyypillistä suora yhdistäminen. Suomen kielessä kaksi substantiiviva voi yleensä muodostaa sanaliiton vain, jos niistä toinen on jossakin muussa sijassa kuin nominatiivissa. Tämän vuoksi nominatiivialkuisia yhdyssanoja voi syntyä sanaliitoista tiivistymällä vain silloin, kun alkuosa ei ole substantiivi (esim. 30). Koska suurin osa uusien yhdyssanojen alkuosista on substantiiveja, suoraan yhdistämisen yleisyys on odotuksenmukaista. Myös muihin sanaluokkiin kuuluvat alkuosat voidaan yhdistää jälkiosaan suoraan (esim. *mobii-limaailma*).

- (30) Feelmaxin *kevytkengät* on tehty tömäkästä materiaalista. Millimetrin paksuinen pohja on sveitsiläistä schöller-keprotecia, jonka sanotaan olevan painoensa nähden viisi kertaa terästä vahvempaa. [TT <- Kauppalehti Presso 17.3.2007, B14.]

Nominatiivialkuisista yhdyssanoista voidaan johtaa uusia yhdyssanoja siinä mistä kaikista muistakin sanoista. Yhdistämällä muodostettuja nominatiivialkuisia uusia yhdyssanoja on kuitenkin niin valtavan paljon, että johtamalla syntyneiden muodosteiden osuus kaikkien nominatiivialkuisten tapausten joukossa jää hyvin pieneksi. Tutkimusotoksessa johtamalla muodostettuja nominatiivialkuisia tapauksia on vain yksi (*sohvaperuna* > *sohvaperunointi*).

Genetiivialkuisista yhdistämällä muodostetuista uusista yhdyssanoista monet voidaan tulkita joko suoraan yhdistetyiksi tai sanaliitoista tiivistyneiksi. Sellaisillekin muodosteille, joiden taustalta ei välttämättä ole tunnistettavissa mitään vakiintunutta sanaliittoa, voidaan helposti mieltää sanaliittomainen välivaihe ennen lopullista yhdistymistä (esim. *parsankuorin*, *toiminnanohjaus*). Selvästi suoraan yhdistettyjä ovat sellaiset genetiivialkuiset yhdyssanat, joiden merkitys ei ole sama kuin niitä vastaavan sanaliiton (esim. *matkustajanpuolikas* ark. halv. 'matkustaja' ≠ *matkustajan puolikas*, *pirunkeppi* 'piikkiaralia' ≠ *pirun keppi*). Myös sanaliitoista tiivistyneiden genetiivialkuisten yhdyssanojen merkitys voi ajan kuluessa eriytyä niitä vastaavasta sanaliitosta. Uusissa sanaliitoista tiivistyneissä yhdyssanoissa semanttisia muutoksia ei kuitenkaan ole yleensä ehtinyt vielä tapahtua (*parinvallinta*, *keittiönkaappi*). Nominatiivialkuisten yhdyssanojen tapaan johtamalla muodostaminen on harvinaista myös genetiivialkuisten sanojen ryhmässä. Johtamalla muodostettuja tapauksia on ryhmässä viisi (esim. *saappaanheittäjä*, *vuodenaikaisuus*).

Muussa sijassa kuin nominatiivissa tai genetiivissä olevan alkuosan sisältäviä yhdyssanoja on koko tutkimusaineistossa vain 28. Näistä 6 on johtamalla muodostettuja. Johdoksista yksi on verbikantainen (*hyväksilukea* > *hyväksiluku*), yksi sanaliitosta johdettu (*hällä väliä* > *hällävälisi*) ja loput substantiiveista johdettuja (esim. *maahantuojia* > *maahantuoda*, *työssäoppiminen* > *työssäoppija*). Yhtä lukuunottamatta kaikki yhdistämällä muodostetut tapaukset ovat sanaliitosta tiivistyneitä, tai

ainakin niiden taustalle on helppo olettaa sanaliittomainen välivaihe⁸² (*puheeksiotto, rännässätallaaja, vettähylkivä*). Poikkeus on suoraan yhdistämällä muodostettu *moninpeli*, jota vastaava sanaliitto ei ole mahdollinen (**monin peli*). Muodosteen alkuosa voidaan tulkita *moni*-sanana instruktiivin sijaan myös adverbiksi (ks. luku 5.3.2). Tällöin voidaan sanoa, että muussa sijassa kuin nominatiivissa ja genetiivissä olevan alkuosan sisältäviä uusia yhdyssanoja syntyy vain johtamisen ja sanaliitoista tiivistymisen kautta. Koska sanaliitoista tiivistymällä syntyvien yhdyssanojen rakenne motivoituu kantana olevan sanaliiton rakenteesta, muulla sijalla alkavat yhdyssanarakenteet eivät näin ollen ole suomen kielessä produktiivisia nimenomaan yhdyssanarakenteina.

6.1.3 Yhteenveto

Suomen kielen uusista yhdyssanoista suurin osa on yhdistämällä muodostettuja. Yhdistäminen on tyypillistä varsinkin yhdyssubstantiiveille sekä nominatiivialkuisille yhdyssanoille. Johtaminen on pääasiallinen muodostuskeino vain yhdysverbien ryhmässä. Niistä noin 75 % on muodostettu johtamalla aiemmista yhdyssanoista. Yhdysverbejä muodostetaan kuitenkin myös yhdistämällä, ja niiden rakennetyyppi on produktiivinen myös yhdistämisen näkökulmasta.

Erilaisista yhdistämisen tavoista käytetyin on suoraan yhdistäminen, joka sekin on tyypillistä erityisesti substantiiveille ja nominatiivialkuisille sanoille. Suoraan yhdistäminen on tyypillistä myös yhdysadjektiiveille, mutta varsinkin *-inen*-loppuisille adjektiiveille on tavallista myös sanaliitoista tiivistyminen sekä johtaminen. Myös yhdistämällä muodostetut yhdysverbit syntyvät yleensä nimenomaan kahdesta sanasta suoraan yhdistämällä.

Sanaliitoista tiivistyminen on tyypillistä genetiivillä ja varsinkin muulla sijalla alkaville yhdyssanoille. Muulla sijalla alkavien yhdyssanojen ryhmässä sanaliitoista tiivistyminen on yksinomainen yhdistämisen keino. Koska aineiston kaikkien muulla sijalla alkavien yhdyssanojen rakenteelle voidaan löytää muu kuin yhdyssanarakenteesta kumpuava motivaatio, muulla sijalla alkavaa yhdyssanarakennetta ei voida pitää suomen kielessä produktiivisena.

6.2 Leksikaalinen monipuolisuus

Tässä luvussa etsin vastausta kysymyksiin siitä, kuinka suuri osa tietyn rakennepiirteen suosiosta on seurausta jonkin yksittäisen sanan toistumisesta ja kuinka monipuolisesti eri rakenteissa käytetään erilaisia leksemejä. Ensimmäisessä alaluvussa (6.2.1) esittelen tarkastelussa käytetyt tunnusluvut. Toisessa alaluvussa (6.2.2) tarkastelen yhdyssanan alkuosien leksikaalista monipuolisuutta ja kolmannessa

82 Aineiston kaikkien yhdistämällä muodostettujen muussa sijassa olevan alkuosan sisältävien yhdyssanojen merkitys on sama kuin vastaavan sanaliiton.

alaluvussa (6.2.3) yhdyssanan jälkiosien leksikaalista monipuolisuutta. Käsitte-
len molemmissa luvuissa leksikaalisen monipuolisuuden yhteyttä kolmeen muut-
tujaan: 1) alkuosan sanaluokkaan (ks. myös 5.2.4), alkuosan muotoon (ks. myös
5.3.1 ja 5.3.2) ja koko muodosteen sanaluokkaan (ks. myös 5.2.2). Luvussa on käy-
tetty osa-aineistoa 3, johon kuuluvat vain 2 leksikaalista osaa sisältävät sanat⁸³ ($n =$
22 144; osa-aineistot on esitelty tarkemmin luvussa 4.3).

Tätä lukua varten olen lemmannut myös aineiston yhdyssanojen alkuosat
perusmuotoihinsa (nominin nominatiiviin ja verbit A-infinitiiviin; taipumattomiin
ja vajaasti taipuviin sanaluokkiin kuuluvien sanojen ja itsenäistä sanaa edustamat-
tomien leksikaalisten elementtien lemmattuna muotona olen käyttänyt ko. haku-
sanassa esiintyvää muotoa). Homonymisten sanojen eri merkitykset on eroteltu
analyysissa vain, jos ero näkyy myös tarkasteltavan muuttujan eri tasoilla. Tämä
tarkoittaa, että esimerkiksi *euro*-alkuosan substantiivinen merkitys 'rahayksikkö'
ja sanaluokaltaan epäspesifi merkitys 'Eurooppaan liittyvä' on laskettu eri leksee-
meiksi tarkasteltaessa alkuosan leksikaalista monipuolisuutta sanaluokan mukaan
mutta samaksi lekseemiksi tarkasteltaessa leksikaalista monipuolisuutta esimer-
kiksi koko muodosteen sanaluokan mukaan. Menettely ei muuta merkittävästi ana-
lyysin tuloksia.

6.2.1 Leksikaalisen monipuolisuuden tarkastelussa käytetyt tunnusluvut

Käytän leksikaalisen monipuolisuuden tarkastelussa apuna kahta toisiaan tukevaa
tunnuslukua: rakenteen sane/lemma-suhdelukua (SLS-luku) sekä rakenteen suh-
teellista entropiaa (H_{rel}). Esittelen tunnusluvut esimerkein valaisten.

Varsinkin lapsenkielen ja käännöskielen tutkimuksessa on perinteisesti käytetty
sanaston monipuolisuuden tarkastelussa sana/sane-suhdelukua (TTR-luku, engl.
type/token-ratio), joka lasketaan jakamalla tekstissä esiintyvien erilaisten sanojen
määrä kaikkien tekstissä esiintyvien saneiden määrällä (tarkemmin esim. Templin
1957: 115; Stubbs 2001: 133–134; Granger 2002; Richards 2009 [1987]). Jarmo
Harri Jantusen (2004: 98–99) mukaan suomen kaltaisissa taivutukseltaan rikkaissa
kielissä on TTR-luvun sijaan parempi käyttää lemma/sane-suhdelukua (LSS-luku),
joka kuvaa tekstissä olevien eri lekseemien määrän (ei TTR-luvun tapaan ainoas-
taan ulkoasultaan erilaisten sanojen määrän) suhdetta tekstin kaikkien saneiden
määrään. Koska tässä tutkimuksessa käytetty aineisto ei ole tavanomainen korpus,
tässä tutkimuksessa lemma/sane-suhdelukua havainnollisempi on sane/lem-
ma-suhde (SLS-luku). Se mahdollistaa ryhmien vertailun yhtä lailla kuin LSS-luku

83 Koko muodosteen sanaluokkajakaumaa käsittelevässä luvussa 5.2.2 on käytetty osa-aineistoa 1, joka sisältää yhden kirjoitusasun kaikista aineiston sanoista. Vain kaksiosaisia yhdyssanoja sisältävä osa-aineisto 3 sopii kuitenkin paremmin myös eri sanaluokkiin kuuluvien yhdyssanojen leksikaalisen monipuolisuuden tarkasteluun, koska siinä pitkän (3 tai useampia leksikaalisia osia) yhdyssanan osana käytetyt vakiintuneet yhdyssanat eivät pääse vääristämään tulosta (esim. aineistossa yleiset *yliopisto*-alkuiset yhdyssanat antaisivat väärän kuvan muutenkin yleisen *yli*-alkuosan käytöstä uudessa sanastossa).

mutta kertoo samalla myös sen, kuinka monen yhdyssanan osana yksi lekseemi aineistossa keskimäärin esiintyy.

SLS-luku lasketaan yksinkertaisesti jakamalla kaikkien tiettyä rakennetta noudattavien yhdyssanojen määrä niissä esiintyvien eri alku- tai jälkiosien määrällä (esim. verbialkuisten yhdyssanojen määrä / yhdyssanan alkuosana käytettyjen eri verbien määrä). Pienin mahdollinen SLS-luku on 1. Mitä pienempi SLS-luku on, sitä monipuolisempi rakenne on leksikaalisesti eli sitä harvemmin samat lekseemit toistuvat rakenteessa. Suuri SLS-luku puolestaan kertoo jakaumasta, jossa monipuolisuutta on vähemmän ja jossa samat lekseemit esiintyvät usein. Suurin mahdollinen SLS-luku on sama kuin ryhmän koko (kuvitteellisessa aineistossa esim. 100 nominatiivialkuista yhdyssanaa / ainoastaan 1 erilainen alkuosalekseemi = SLS = 100). SLS-arvoja vertailemalla voidaan tunnistaa rakenteet, joissa esiintyy erityisen paljon tai vähän eri lekseemejä.

Sekä TTR-, LLS- että SLS-arvot ovat riippuvaisia tekstin pituudesta (tai tämän tutkimuksen tapauksessa aineiston koosta; TTR ja LSS ovat sitä pienempiä, mitä pidempiä tekstit ovat, SLS puolestaan sitä suurempi, mitä pidempi teksti on). Ilmiö johtuu siitä, että vaikka pidemmässä tekstissä on enemmän saneita kuin lyhyessä, eri sanamuotojen tai lemموjen määrä ei kasva samassa suhteessa, sillä samat sanat toistuvat tekstissä. Sanaston monipuolisuutta kuvaavatkin paremmin standardoidut suhdeluvut (STTR, SLLS ja SSLS), jotka saadaan laskemalla samanpituisten tekstinosien (esim. 1 000 sanaa) suhdelukujen keskiarvo. (Esim. Cross 1977; Stubbs 2001: 133; Richards 2009 [1987]; McCarthy–Jarvis 2010: 381.) Omassa tutkimuksessani hyvin erikokoiset ryhmät vaikeuttavat standardointia: osa ryhmistä on hyvin pieniä, mutta koska pienet ryhmät antavat ylipäänsä suuria epäluotettavampia tuloksia, suuria ryhmiä ei ole järkevää standardoida kovin pieniksi. Tämän vuoksi olen päättänyt laskemaan SSLS-arvon suurimmista ryhmistä kahdesti – sekä suurista otoksista (600 yhdyssanaa; koko valittu siten, että mahdollisimman monesta ryhmästä saadaan tarvittavan kokoinen suuresta joukosta laskettu SSLS-arvo) sekä pienistä otoksista (100 yhdyssanaa; koko valittu siten, että mahdollisimman monista ryhmästä saadaan keskenään vertailukelpoiset SSLS-arvot) (ks. myös J. H. Jantunen 2004: 99). SSLS-arvojen laskemiseksi olen ottanut kustakin ryhmästä 10 satunnaisotosta ja laskenut niiden keskiarvon. Mikäli ryhmään kuuluu alle 100 yhdyssanaa, olen standardoidun arvon sijaan laskenut ainoastaan standardoimattoman SLS-arvon koko ryhmästä.

Pelkät SLS-arvot eivät kerro mitään jakaumien muodosta. Niiden avulla ei voida havaita esimerkiksi sellaisia ryhmiä, joissa muut lekseemit esiintyvät harvoin ja korkea SLS-arvo on seurausta vaikka vain yhden tai kahden lekseemin jatkuvasta toistumisesta. Tällainen joukko on esimerkiksi jo aiemmin esiin nostamani pronominalkuisten yhdyssanojen ryhmä, jossa jopa 71,3 % sanoista alkaa *moni*-kvanttoripronominilla (luku 5.2.4). Tämän vuoksi myös ryhmän SLS-arvot ovat korkeita (SSLS/100 = 7,2). Myös kotoisella prefiksinkaltaisella elementillä alkavien yhdyssanojen SLS-arvot ovat muihin ryhmiin verrattuna melko korkeita (SSLS/100 = 2,6). Oleellinen ero pronominilla ja kotoisella prefiksinkaltaisella elementillä alkavien

yhdyssanojen välillä on se, että prefiksinkaltaisten alkuosien ryhmässä eri lekseemit esiintyvät huomattavasti tasaisemmin, eikä niiden joukossa mikään alkuosa nouse *moni*-pronominin tapaan ylitse muiden. Siksi – kuten yleensä aina keskimääräisten arvojen tarkastelun lisäksi – leksikaalisen monipuolisuuden tarkastelussa on syytä huomioida myös jakauman hajonta.

Suhteellinen entropia (H_{rel}) kuvaa kategorisen muuttujan hajontaa eli sitä, kuinka tasaisesti havainnot ovat jakautuneet eri luokkiin. Menetelmä antaa välillä 0–1 olevan tunnusluvun. Mitä lähempänä arvo on yhtä, sitä tasaisemmin arvot ovat jakautuneet (ts. sitä vaikeampaa on ennustaa tietyn arvon esiintymistä jakaumassa). Suhteellinen entropia on 1 jakaumassa, jossa on yhtä paljon kaikkia arvoja. Lähempänä nollaa oleva arvo puolestaan kertoo jakauman epätasaisuudesta. Suhteellinen entropia on 0 jakaumassa, jossa ei ole lainkaan variaatiota vaan kaikki muuttujan arvot ovat samoja (ts. tietyn arvon esiintyminen jakaumassa voidaan ennustaa täysin varmasti). Suhteellinen entropia auttaa tunnistamaan jakaumat, joissa jotkin sanat tuottavat erityisen suuren tai pienen osan kaikista rakennetta noudattavista tapauksista (esim. Gries 2009: 112–113, 2010: 272–273)⁸⁴.

Stefan Th. Griesin (2010: 273) mukaan suhteellinen entropia mahdollistaa erikokoisten ryhmien vertailun. Niklas Wilming ym. (2011: 16–17) kuitenkin huomauttavat, että myös suhteellinen entropia on jossain määrin riippuvainen otoksen koosta. Otoksien standardoinnin tärkeys on ilmeistä myös tämän luvun analyysiosassa (erityisen helposti otoksien vaikutuksen voi nähdä jälkiosia käsittelevästä luvusta 6.2.3). Vertailukelpoisten tulosten saamiseksi olen käyttänyt myös suhteellisten entropioiden laskemisessa suurista ryhmistä otettuja samankokoisia satunnaisotoksia (standardoidut suhteelliset entropiat $SH_{rel}/600$ yhdyssanaa ja $SH_{rel}/100$ yhdyssanaa; mikäli ryhmään kuuluu alle 100 yhdyssanaa, olen laskenut ainoastaan standardoimattoman H_{rel} -arvon koko ryhmästä). SH_{rel} -arvojen laskemiseksi olen ottanut kustakin ryhmästä 10 satunnaisotosta ja laskenut niiden keskiarvon.

Kaiken kaikkiaan yhdyssanojen jälkiosat ovat leksikaalisesti hieman alkuosia monipuolisempia (koko osa-aineistossa [$n = 22\ 144$] alkuosien $SLS = 4$ ja $H_{rel} = 0,91$; jälkiosien $SLS = 3,2$ ja $H_{rel} = 0,93$ ⁸⁵). Toisin sanoen jälkiosien joukossa on useampia eri lekseemejä kuin alkuosien. Ero on todennäköisesti seurausta yhdyssanan määrite- ja edusosan erilaisista tehtävistä. Suomen kielessä enemmistönä olevissa määritisyhdyssanoissa jälkimmäinen osa on muodosteen edusosa, joka osoittaa koko muodosteen tarkoitteen ja jolla on tyypillisesti itsenäinen semanttinen merkitys

84 Muita tapoja kategorisen muuttujan variaation tarkasteluun esittelevät esimerkiksi Cichocki ja Amari 2010. Samantapaisella periaatteella toimii myös sosiologiassa käytetty IQV eli *index of qualitative variation* (Mueller–Schuessler 1961).

85 Alku- ja jälkiosista esitetyt tunnusluvut ovat samankokoisista aineistoista ja siten keskenään vertailukelpoiset. Mikäli tunnuslukuja tahdotaan verrata tämän luvun alaluvuissa esiteltyihin ryhmäkohtaisiin tunnuslukuihin, myös alku- ja jälkiosien hajontaa kuvaavat tunnusluvut on laskettava pienempien satunnaisotosten keskiarvosta: Alkuosien $SSL/600 = 1,2$ ja $SH_{rel}/600$ yhdyssanaa = 0,98. Jälkiosien $SSL/600$ yhdyssanaa = 1,1 ja $SH_{rel}/600$ yhdyssanaa = 0,99. Sekä alku- että jälkiosien $SSL/100 = 1$ ja $SH_{rel}/100 = 1$.

(esim. *avopizza* '(erotuksena umpipizzasta) pizza, jonka toista puolta ei ole käännetty toisen päälle kanneksi'; *summayhdyssanassa* yhdysosat ovat semanttisesti yhdenvertaiset ja yleensä molemmilla osilla on itsenäinen semanttinen merkitys, esim. *lusikka-haarukka* 'ruokailuväline, jonka toinen pää on lusikka ja toinen haarukka'). Määritysyhdyssanan alkuosa puolestaan määrittää jälkiosaa tavalla tai toisella. Sisältösanojen lisäksi määrittäviin tehtäviin sopivat myös sanat, joilla ei ole omaa itsenäistä semanttista merkitystä (esim. intensifioivat tai kieltoa ilmaisevat prefiksinkaltaiset alkuosat kuten sanoissa *superlehmä* 'erittäin paljon lypsävä lehmä' ja *epäsoitin* 'soittamiseen käytettävä esine, jota ei ole alun perin tehty soittimeksi'). Koska (suomen) kielessä on enemmän sisältösanoina kuin muotosanoja (termeistä ks. TTP: Kielitiede s. v. *sisältösana* ja *muotosana*), on luonnollista, että yhdyssanan jälkiosat ovat leksikaalisesti jonkin verran monipuolisempia kuin alkuosat. Sisältösanat ovat kuitenkin tavallisia myös yhdyssanan alkuosina (esim. *jäägolf*), eikä eroa alku- ja jälkiosien monipuolisuudessa olekaan paljoa.

6.2.2 Alkuosa

Taulukossa 15 esitellään tiedot tutkimusaineiston yhdyssanojen alkuosan sane/lemma-suhteesta ([S]SLS) ja suhteellisesta entropiasta ($[S]H_{rel}$) alkuosan sanaluokan mukaan jaoteltuna. Lisäksi mukana ovat tiedot kaikkien ryhmään kuuluvien yhdyssanojen määrästä ja rakenteessa esiintyvien eri alkuosien määrästä sekä yleisimmistä alkuosalekseemeistä.

TAULUKKO 15. Tutkimusaineiston yhdyssanojen alkuosan leksikaalinen monipuolisuus alkuosan sanaluokan mukaan

Alkuosan sanaluokka	SSLS /100	SSLS /600	SLS	SH _{rel} /100	SH _{rel} /600	H _{rel}	Ryhmän yhdyssanat/eri alkuosaleksemit	Yleisimmät alkuosaleksemit
Substantiivi	1	1,2	3,7	1	0,99	0,91	17 680/4 795	<i>netti</i> (144 kpl) <i>verkko</i> (122 kpl)
Sana- luokaltaan epäspesifi	1,8	4	7,6	0,95	0,89	0,82	2 052/269	<i>bio-</i> (108 kpl) <i>eko-</i> (85 kpl)
Adjektiivi	1,8	3,9	5,9	0,94	0,87	0,84	1 238/210	<i>uusi</i> (91 kpl) <i>luomu</i> (55 kpl)
Useaa sanaluokkaa edustava	1,5	3,4	3,7	0,96	0,9	0,89	714/193	<i>viher-</i> (41 kpl) <i>itse</i> (27 kpl)
Taipumaton tai vajaasti taipuva sanaluokka	1,8	-	2,1	0,94	-	0,92	153/73	<i>läpi</i> (11 kpl) <i>poikki</i> (11 kpl)
Verbi	1,7	-	1,8	0,94	-	0,94	108/61	<i>uida</i> (10 kpl) <i>syntyä</i> (7 kpl)
Pronomini	7,2	-	7,2	0,48	-	0,48	101/14	<i>moni</i> (72 kpl) <i>minä</i> (8 kpl)
Numeraali		-	4,1		-	0,78	90/22	<i>nolla</i> (25 kpl) <i>kaksi</i> (17 kpl)
Partisiippi	-	-	1	-	-	1	8/8	<i>kapeneva</i> (1 kpl) <i>kuollut</i> (1 kpl) + muita 1 krt esiintyviä

Kuten taulukosta 15 nähdään, alkuosien leksikaalinen monipuolisuus vaihtelee paljon alkuosan sanaluokan mukaan. Suurin standardoitu SSLS/100-arvo on pronominalkuisilla yhdyssanoilla (SSLS/100 = 7.2). Seuraavaksi suurin SLS-arvo on numeraaleilla (standardoimaton SLS = 4,1; tapauksia 90, joten arvoa voidaan pitää kutakuinkin vertailukelpoisena SSLS/100 arvon kanssa). Näissä ryhmissä samat alkuosat toistuvat siis keskimääräin useammin kuin muissa ryhmissä. Tämä on sikäli odotuksenmukaista, että pronominiinien ja (yhden leksikaalisen osan sisältävien) numeraalien ryhmät ovat kielessä kaiken kaikkiaan melko pieniä. Sanaluokkien koko ei kuitenkaan selitä kaikkea taulukossa näkyvää vaihtelua, sillä myös kielessä hyvin erikokoisten sanaluokkien SLS-arvot voivat olla samansuuntaisia: esimerkiksi useaa sanaluokkaa edustavien sanojen, adjektiivien sekä sanaluokaltaan epäspesifien SSLS/600-arvot ovat kaikki välillä 3,4–4 ja useaa sanaluokkaa edustavien sanojen, adjektiivien, sanaluokaltaan epäspesifien, taipumattomien ja vajaasti taipuvien sanaluokkien sekä verbien SSLS/100-arvot välillä 1,5–1,8. Yhdyssanan

alkuosana selvästi monipuolisimmin käytettävä sanaluokka on suomen kielen suurin sanaluokka substantiivit ($SLS/100 = 1$, $SLS/600 = 1,2$, $SLS/100 = 1$).

Taulukossa 15 olevien H_{rel} -arvojen perusteella voidaan päätellä, että eri sanaluokkiin kuuluvien yhdyssanan alkuosien leksikaaliset jakaumat ovat epätasaisimpia pronomini- ja numeraalialkuisissa yhdyssanoissa (pronominiin $SH_{rel}/100 = 0,48$, numeraalien $H_{rel} = 0,78$). Tämä merkitsee, että ryhmissä ei ole ainoastaan vähän eri alkuosia, kuten SLS-arvot osoittavat, vaan ryhmissä käytetään joitakin alkuosaleksemejä huomattavasti enemmän kuin toisia. Muissa sanaluokissa jakaumat ovat selvästi tasaisempia ($SH_{rel}/100$ -arvot välillä $0,94-1$, $SH_{rel}/600$ -arvot välillä $0,87-0,99$). Adjektiivien, taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvien alkuosien, useaa sanaluokkaa edustavien alkuosien ja sanaluokaltaan epäspesifien alkuosien jakaumat ovat myös H_{rel} -arvojen perusteella keskenään sanantyyppisiä. Eri sanaluokkia edustavista alkuosista tasaisin jakauma on substantiiveilla ($SH_{rel}/600 = 0,99$ ja $SH_{rel}/100 = 1$).

Pronominiin ryhmässä sekä jakauman epätasaisuus (H_{rel}) että samojen alkuosien korkea esiintymisfrekvenssi (SLS) selittyvät *moni*-kvantoripronominin yleisyydellä. Se on alkuosana jopa 71,3 %:ssa ryhmän tapauksista (72/101 kpl, ks. myös luku 5.2.4). Seuraavaksi yleisin pronominalalkuosa on kahdeksassa sanassa esiintyvä *minä* (lisäksi muodot *mie* ja *mää* esiintyvät kumpikin yhdessä sanassa). Kaikkiaan aineistossa esiintyy yhdyssanan alkuosana 14 eri pronominia. Aineisto osoittaa, että muulla kuin *moni*-pronominilla alkavat yhdyssanat ovat suomen uudessa yhdyssanastossa harvinaisia (aineistossa 29 kpl) ja että pronomineista vain osaa käytetään produktiivisesti yhdyssanojen alkuosana.

Numeraalialkuosien leksikaalinen yksipuolisuus selittyy osaksi yksisanaisten numeraalien vähyydellä suomen kielessä (koska tässä luvussa käytettyyn osa-aineistoon kuuluu vain kaksi leksikaalista osaa sisältäviä yhdyssanoja, yhdyssanan määriteosana toimivat yhdysnumeraalit eivät näy aineistossa). Yksisanaistenkin numeraalialkuosien jakauma on kuitenkin epätasainen. Yleisimmät numeraalialkuosat aineistossa ovat *nolla* (25 kpl), *kaksi* (17 kpl), *yksi* (11 kpl), *kolme* (5 kpl) ja *viisi* (5 kpl). Kaikkiaan aineistossa esiintyy yhdyssanan alkuosana 22 eri numeraalia (numeraalialkuisia yhdyssanoja yht. 90 kpl; kirjaimin ja numeroin kirjoitetut numeraalit on laskettu eri numeraalin esiintymiksi). Lähes kaikki yksisanaiset perusluvut esiintyvät aineistossa yhdyssanan alkuosana. Lisäksi mukana on sentyyppisiä ilmauksia kuin *tusina* ja *neljännes*.

Suuri osa taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvista sanoista on muotosanoja, joita on kielessä kaiken kaikkiaan vähemmän kuin sisältösanoja. Ryhmän yleisimmät alkuosat ovat *läpi* (11 kpl), *poikki* (11 kpl), *myötä* (10 kpl), *kanssa* (6 kpl) ja *pois* (5 kpl). Kaikkiaan aineistossa esiintyy yhdyssanan alkuosana 73 eri taipumattomiin tai vajaasti taipuviin sanaluokkiin kuuluvaa sanaa (taipumatomalla tai vajaasti taipuvalla sanalla alkavia yhdyssanoja yht. 153 kpl). Ryhmän leksikaalista monipuolisuutta lisää se, että samasta vartalosta muodostetut vajaasti taipuvat sanat on katsottu kukin omaksi lekseemikseen (tämä on yleensä myös sanakirjojen käytäntö, ks. myös Koivisto 2013: 40–41). Mikäli myös vajaasti taipuvat

sanat yhdistetään saman vartalon alle, ryhmän sane/lemma-suhde ja jakauman muoto pysyvät lähes samoina (SSLS/100 nousee 1,8:sta 2:een ja SH_{rel}/100 laskee 0,94:stä 0,92:een). Eri vartaloita on tällöin käytössä 63. Kolme yleisintä sanaa pysyvät samoina, mutta neljänneksi nousevat *yl-*vartalon sisältävät sanat (*yl*⁸⁶ ja *ylös* yht. 8 kpl) ja viidenneksi *al-*vartalon sisältävät sanat (*alas*, *alle* ja *alta* yht. 7 kpl).

Useaan sanaluokkaan kuuluvien alkuosien ryhmän yleisimmät alkuosat ovat *viher-* (41 kpl), *itse* (27 kpl), *puna[inen]* (23 kpl), *homo* (18 kpl), *tanssi[a]* (17 kpl) ja *kulta* (16 kpl). Kaikkiaan aineistossa esiintyy yhdyssanan alkuosana 193 eri useaan sanaluokkaan kuuluvaa sanaa (useaan sanaluokkaan kuuluvalla sanalla alkavia yhdyssanoja yht. 714 kpl). Valtaosa ryhmän alkuosista on substantiivi-adjektiiveja. Looginen selitys ryhmän leksikaaliselle yksipuolisuudelle esimerkiksi substantiivialkuosiin verrattuna onkin sekä substantiiveina että adjektiiveina tyypillisesti käytettävien sanojen keskittyminen tiettyihin aihepiireihin. Substantiivi-adjektiivit ovat tavallisesti joko ihmiseen viittaavia sanoja tai värinimityksiä, joita on kielessä tietysti kaiken kaikkiaan vähemmän kuin muita substantiiveja ja adjektiiveja (ks. myös luku 5.2).

Adjektiivialkuosien ryhmässä on monia usein toistuvia alkuosia (esim. *uusi* 91 kpl, *luomu* 55 kpl, *virtuaalinen* 52 kpl, *kylmä* 44 kpl, *mobiili* 44 kpl ja *kevyt* 41 kpl). Kaikkiaan aineistossa esiintyy yhdyssanan alkuosana 210 eri adjektiivia (adjektiivialkuisia yhdyssanoja yht. 1 238). Alkuosissa toistuvia sanoja ei voida kutsua muotilauiksi samassa mielessä kuin yhdyssanojen jälkiosissa usein toistuvia adjektiiveja nimitetään muotilopuiksi (ks. luku 5.2.2), vaan ne ovat ennemmin ajattomia ja aina tarpeellisia yhdyssanan alkuosia tai heijastavat syntyaikanaan yhteiskunnassa pinnalla olleita aihepiirejä.

Sanaluokaltaan epäspesifien alkuosien ryhmässä useimmin toistuvat elementit ovat prefiksinkaltaisia epäitsenäisiä elementtejä (*bio-* 108 kpl, *eko-* 85 kpl, *nano-* 68 kpl, *perus-* 66 kpl, *pika-* 62 kpl, *euro-* 'Eurooppaan liittyvä' 59 kpl ja *täsmä-* 59 kpl). Kaikkiaan aineistossa esiintyy yhdyssanan alkuosana 269 eri sanaluokaltaan epäspesifiä sanaa tai elementtiä (sanaluokaltaan epäspesifillä alkuosalla alkavia yhdyssanoja yht. 2 052 kpl). Vaikka useimmin esiintyvät elementit ovat tässä ryhmässä vielä yleisempiä kuin adjektiivialkuosien ryhmässä, sanaluokaltaan epäspesifien ryhmä on kuitenkin kaiken kaikkiaan leksikaalisesti hiukan monipuolisempi.

Yhdyssanojen substantiivialkuosien leksikaalinen monipuolisuus ja jakauman tasaisuus selittyy laajalla valinnanvaralla: suurin osa suomen kielen sanoista on substantiiveja. Kuitenkin myös substantiivialkuosista jotkin toistuvat toisia useammin. Aineistossa yleisimmät alkuosat ovat aikakauttaan heijastavat *netti* (144 kpl)

86 Esimerkiksi Häkkisen (1987: 19) mukaan *yli-* ja *ali-* ovat produktiivisia yhdyssanan adverbialkuosia. Tässä tutkimuksessa *yli-* ja *ali-* alkuosat on luokiteltu adverbeiksi vain silloin, kuin niillä viitataan ylitse tai alitse liikkumiseen (myös ajassa, esim. *yliyö*). Muussa tapauksessa *yli-* ja *ali-* alkuosat on katsottu epäitsenäisiksi prefiksinkaltaisiksi elementeiksi (ks. myös ISK § 172 ja 430; luokittelussa käytetyistä periaatteista ks. myös luku 5.3.2.2). Tutkimusaineistossa on vain 4 adverbina käytettyä yhdyssanan *yli-* alkuosaa (esim. *ylimaalaus*, *ylikirjoittaa*; prefiksinkaltaista käyttöä esim. *ylieläke* 'liian suuri eläke', *ylivahva* 'liian vahva'). *Ali-* alkuosaa ei ole käytetty tutkimusaineistossa lainkaan adverbina (prefiksinkaltaista käyttöä esim. *alikäyttö* 'liian vähäinen käyttö', *alivilkkaus* 'liian vähäinen vilkkaus').

verkko (122 kpl), *euro* ('valuuttayksikkö' 111 kpl), *media* (95 kpl), *vesi* (94 kpl) ja *turva* (85 kpl). Kaikkiaan aineistossa esiintyy 4 795 eri substantiivialkuosaa (substansiivialkuisia yhdyssanoja yht. 17 680 kpl).

Verbit esiintyvät yhdyssanan alkuosina vain verbivartaloina sekä *-mA-* ja *-in-*loppuisissa muodoissa. Verbaalisia merkityksiä ilmaistaan kuitenkin yhdyssanoissa myös substantiivialkuosilla (ks. tarkemmin luku 5.3.2.3), ja varsinaiset verbimuodot esiintyvätkin yhdyssanan alkuosina melko harvoin. Tutkimusaineistossa yleisimmät verbialkuosat ovat *uida* (10 kpl), *syntyä* (7 kpl), *asua* (5 kpl), *elää* (5 kpl) ja *istua* (4 kpl). Kaikkiaan aineistossa esiintyy yhdyssanan alkuosana 61 eri verbiä (verbialkuisia yhdyssanoja yht. 108 kpl). Eri partisiippialkuosia on aineistossa vain 8, kuten partisiippialkuisia yhdyssanojakin. Näin pienen aineiston perusteella ei voida tehdä päätelmiä yhdyssanan partisiippialkuosien leksikaalisesta monipuolisuudesta koko populaatiossa.

Yllä olen käsitellyt yhdyssanojen alkuosan leksikaalista monipuolisuutta alkuosan sanaluokan mukaan. Seuraavaksi siirryn käsittelemään alkuosan leksikaalista monipuolisuutta alkuosan muodon mukaan. Alkuosan sanaluokka on jossain määrin yhteydessä alkuosan muotoon (ks. tarkemmin luku 7.2.1). Alkuosan muodon yhteydet rakenteen leksikaaliseen monipuolisuuteen eivät kuitenkaan ole yhtä selviä kuin alkuosan sanaluokan. Taulukkoon 16 on koottu tiedot tutkimusaineiston yhdyssanojen alkuosan sane/lemma-suhteesta ([S]SLS) ja suhteellisesta entropiasta ([S]H_{rel}) alkuosan muodon mukaan ryhmiteltynä. Lisäksi mukana ovat tiedot kaikkien ryhmään kuuluvien yhdyssanojen määrästä ja rakenteessa esiintyvien eri alkuosien määrästä sekä yleisimmistä alkuosalekseemeistä. Jaottelu noudattaa luvussa 5.3.1 käytettyä alkuosan muodon karkeaa jakoa.

TAULUKKO 16. Tutkimusaineiston yhdyssanojen alkuosan leksikaalinen monipuolisuus alkuosan muodon mukaan

Alkuosan muoto	SSLS /100	SSLS /600	SLS	SH _{rel} /100	SH _{rel} /600	H _{rel}	Ryhmän yhdyssanat/ eri alkuosalekseemit	Yleisimmät alkuosalekseemit
Nominatiivi	1	1,2	3,7	1	0,99	0,92	17 599/4 732	<i>netti</i> (144 kpl) <i>verkko</i> (121 kpl)
Epäitsenäinen elementti	1,8	4	7,6	0,95	0,89	0,82	2 052/269	<i>bio-</i> (108 kpl) <i>eko-</i> (85 kpl)
Yhdysosamuoto	1,4	2,4	3,4	0,96	0,89	0,86	1 359/400	<i>uus-</i> (84 kpl) <i>virtuaali-</i> (52 kpl)
Genetiivi	1,2	1,5	1,7	0,98	0,95	0,94	865/521	<i>maailma</i> (30 kpl) <i>elämä</i> (25 kpl)
Taipumaton tai vaillinaisesti taipuva sana	2,1	-	3	0,89	-	0,84	241/81	<i>pikku</i> (39 kpl) <i>ensi</i> (26 kpl)
Muu sijamuoto	-	-	1,5	-	-	0,9	28/19	<i>maa</i> (6 kpl) <i>työ</i> (5 kpl)

Kuten taulukosta 16 nähdään, alkuosan muodolla on alkuosan sanaluokkaa vähemmän vaikutusta alkuosien leksikaaliseen monipuolisuuteen (vrt. edellinen taulukko 15, s. 163). Useimmin samoja alkuosaleksemejä esiintyy taipumattomien ja vaillinaisesti taipuvien alkuosien ($SSLS/100 = 2,1$) sekä epäitsenäisten elementtien ($SSLS/600 = 4$) ryhmissä. Epäitsenäisten elementtien ryhmässä huomiota kiinnittää myös koko ryhmän korkea SLS-luku (7,6). Luku on ryhmän kokoon (2 052 yhdyssanaa; vrt. esim. huomattavasti suuremman nominatiivialkuisten ryhmän $SLS = 3,7$) nähden suuri ja viestii siitä, että vaikka epäitsenäisellä elementillä alkavien yhdyssanojen otosta kasvatetaan, mukaan ei juuri tule uusia erilaisia alkuosia. Seuraavaksi vähiten monipuolisuutta on yhdysosamuotojen ($SSLS/600 = 2,4$) ja genetiivialkuosien ryhmissä ($SSLS/600 = 1,5$). Ryhmistä leksikaalisesti monipuolisin on suurin eli nominatiivialkuosien ryhmä ($SSLS/600 = 1,2$). Erot nominatiivi- ja genetiivialkuisten välillä ovat kuitenkin pieniä. Molempiin ryhmiin mahtuu laaja kirjo sisältösanoja, joista suurin osa on yhdyssanojen alkuosissa monipuolisesti esiintyviä substantiiveja.

Eri lekseemit ovat jakautuneet melko tasaisesti kaikissa alkuosan muotoryhmissä ($SH_{rel}/100 = 0,89-1$ ja $SH_{rel}/600 = 0,89-0,99$). Tässä suhteessa muista eniten poikkeavat taipumattomat ja vaillinaisesti taipuvat alkuosat ($SH_{rel}/100 = 0,89$). Seuraavaksi epätasaisin jakauma on epäitsenäisillä elementeillä ja yhdysosamuodoilla (epäitsenäisten elementtien $SH_{rel}/100 = 0,95$ ja $SH_{rel}/600 = 0,89$; yhdysosamuotojen $SH_{rel}/100 = 0,96$ ja $SH_{rel}/600 = 0,89$). Tasaisimpia ovat genetiivi- ($SH_{rel}/600 = 0,95$) ja nominatiivialkuosien jakaumat ($SH_{rel}/600 = 0,99$). Kaikkien alkuosamuotojen jakauma on selvästi tasaisempi kuin esimerkiksi pronomini- ja numeraalialkuosien (vrt. taulukko 15 edellä, s. 163).

Samat alkuosat esiintyvät yhdyssanan alkuosan muotoryhmistä tiheimmin taipumattomien ja vajaasti taipuvien sanojen sekä epäitsenäisten elementtien ryhmissä. Epäitsenäisten elementtien ryhmästä suurimman osan muodostavat kotoiset ja vierasperäiset prefiksinkaltaiset elementit (ks. tarkemmin luku 5.3.2.2). Alkuperältään erilaisten prefiksinkaltaisten alkuosien leksikaalinen variaatio eroaa toisistaan. Kotoisia prefiksinkaltaisia elementtejä on kaiken kaikkiaan olemassa vähemmän kuin vierasperäisiä, ja kotoisten prefiksinkaltaisten elementtien ryhmässä onkin vähemmän vaihtelua kuin vierasperäisten ryhmässä (kotoisella prefiksinkaltaisella elementillä alkavien $SSLS/600 = 10$ ja $SSLS/100 = 2,6$, ryhmässä yhdyssanoja yht. 998 kpl / eri alkuosia 63 kpl; vierasperäisellä prefiksinkaltaisella elementillä alkavien $SSLS/600 = 5,6$ ja $SSLS/100 = 2,5$, ryhmässä yhdyssanoja yht. 922 / eri alkuosia 131 kpl). Tästä huolimatta aineiston kaikkein yleisimmät prefiksinkaltaiset elementit ovat vierasperäisiä. Kotoisista prefiksinkaltaisista elementeistä aineistossa useimmin esiintyvät *perus-* (66 kpl), *pika-* (62 kpl), *täsmä-* (59 kpl), *etä-* (52 kpl) ja *yli-* (51 kpl). Vierasperäisistä useimmin esiintyvät *bio-* (108 kpl), *eko-* (85 kpl), *nano-* (68 kpl), *euro-* (59 kpl), *digi-* (57 kpl) ja *super-* (51 kpl). Erittäin usein toistuvien alkuosien lisäksi vierasperäisten elementtien ryhmässä on myös useita vain kerran esiintyviä alkuosia (esim. *betaäiti*, *pikolitra*, *polyrottinki*, *subteksti*, *yot-tatavu*), kun taas kotoisista elementeistä yhden kerran esiintyvät vain muodosteessa

intensifioivana käytetty *homo*-⁸⁷ (*homohyvä* 'todella hyvä, sikahyvä') ja merkityksessä 'mieli- suosikki-' esiintyvä *lempi-* (*lempivaate*). Myös suhteellisen entropian arvot osoittavat, että eri lekseemien määrät ovat jakautuneet huomattavasti tasaisemmin kotoisten ($SH_{rel}/600 = 0,91$, $SH_{rel}/100 = 0,94$) kuin vierasperäisten elementtien ryhmässä ($SH_{rel}/600 = 0,78$, $SH_{rel}/100 = 0,88$).

Myös yhdysosamuotojen kohdalla on syytä kiinnittää huomiota alaryhmien (ks. luokittelu luvusta 5.3.2.3) keskinäisiin eroihin. Alaryhmistä joissakin on hyvin vähän vaihtelua. Esimerkiksi muiden⁸⁸ konsonanttivartaloiden ryhmässä sane/lemma-suhde on varsin korkea ($SSLS/100 = 14,3$). Yhteensä ryhmään kuuluu 151 yhdysosanaa, joissa esiintyy vain 7 eri alkuosaa. Näiden seitsemän alkuosan jakauma on lisäksi epätasainen (ryhmän $SH_{rel}/100 = 0,7$). Ryhmän yleisin alkuosa on 84 kertaa esiintyvä *uus-*vartalo. Seuraavina tulevat *pien-* (23 kpl), *täys-* (21 kpl) ja *suur-*vartalot (18 kpl). *Tois-* (2 kpl), *pers-* (2 kpl) ja *nuor-*vartaloiden (1 kpl) esiintymät ovat harvinaisia. Kaiken kaikkiaan suomen kielessä on vain vähän sanoja, joilla on kielopin suomat mahdollisuudet esiintyä yhdyssanan alkuosana konsonanttivartalona (ISK [§ 416] mainitsee tutkimusaineistossani esiintyvien lisäksi myös alkuosat *enkel-* ja *jumal-*). Voidaan siis sanoa, että rakenne tuottaa uusia muodosteita hyvin produktiivisesti siihen nähden, kuinka paljon kielessä on rakenteeseen sopivia sanoja (ks. myös Aronoff 1976; Bauer 2001: 5–7). Koska rakenteeseen sopivia sanoja on kuitenkin hyvin vähän, on myös syytä pohtia, voiko rakennetta sinänsä pitää yhdyssananmuodostuksessa produktiivisena vai onko enemmän puhuttava tiettyjen alkuosalekseemien produktiivisuudesta.

Muulla konsonanttivartalolla alkavia yhdyssanoja voidaan verrata esimerkiksi yhdyssanoihin, joiden alkuosa on *-inen*-loppuisista vierassanoista käytettävä *-i*-loppuinen yhdysosamuoto. Myös tämä joukko on leksikaalisesti melko yksipuolinen: 165 yhdyssanan ryhmässä esiintyy vain 30 alkuosaa ($SSLS/100 = 4,2$; $SH_{rel}/100 = 0,77$). Yleisin alkuosa on *virtuaali-* (52 kpl), jota seuraavat selvästi harvinaisemmat *sosiaali-* (26 kpl) ja *digitaali-* (16 kpl). Leksikaalisesta yksipuolisuudesta huolimatta nimenomaan rakennetta – ei yksittäisiä lekseemejä – voidaan pitää produktiivisena, sillä *-i*-loppuisia yhdysosamuotoja muodostetaan mekaanisesti *-inen*-loppuisista vierassanoista. Suomen kieleen myös saadaan uusia tähän rakenteeseen sopivia alkuosia sitä mukaa, kun uusia vieraskantaisia *-inen*-loppuisia sanoja muodostetaan.

Yhdysosamuotojen alaryhmistä useat ovat niin pieniä, että niiden leksikaalisesta monipuolisuudesta ei voida tämän tutkimuksen aineiston perusteella sanoa mitään varmaa. Lähimpänä koko ryhmän yhteisiä arvoja on alaryhmistä suurin: *-nen*-loppuisten sanojen *-s*-loppuisten yhdysosamuotojen $SSLS/600 = 2,7$ (koko ryhmällä 2.4) ja $SH_{rel}/600 = 0,89$ (myös koko ryhmällä 0,89). Ryhmään kuuluu

87 *Homo* on tietysti alkuun lainasana, mutta tässä intensifioivassa merkityksessä katson sen esimerkiksi *sika-* tai *tos-*elementtejä vastaavaksi kotoiseksi prefiksiksi, sillä *homo*-elementin prefiksinkaltainen käyttö on alkanut vasta suomen kielessä eikä elementtiä ole alkuun lainattu prefiksinkaltaiseen käyttöön.

88 *Muu* viittaa tässä muihin kuin *-nen*-loppuisten sanojen yhdysosamuotoina toimiviin ja niiden konsonanttivartalon kanssa identtisiin *-s*-loppuisiin muotoihin (ks. luokittelu luvusta 5.3.2.3).

yhteensä 780 yhdyssanaa, joissa esiintyy 255 eri alkuosaa. Yleisin alkuosa on *nais-* (46 kpl; *nais-*määriteosasta ks. lisää Tyysteri 2011c). Seuraavina tulevat *ihmis-* (28 kpl), *kansalais-* (26 kpl), *sijais-* (25 kpl) ja *vertais-* (23 kpl).

Eri sijamuotoja edustavien alkuosien leksikaalisessa monipuolisuudessa on vain pieniä eroja. Niistä leksikaalisesti monipuolisin on nominatiivimuotoisten alkuosien joukko ($SLS/600 = 1,2$, $SH_{rel}/600 = 0,99$). Yhteensä ryhmässä on 17 599 yhdyssanaa, joissa esiintyy 4 732 eri alkuosaa. Yleisimmät alkuosat ovat *netti* (144 kpl) ja *verkko* (121 kpl), joita seuraavat *euro* ('rahayksikkö', 110 kpl), *media* (94 kpl) ja *vesi* (89 kpl). Leksikaalisesti toiseksi monipuolisin on genetiivialkuosien joukko ($SLS/600 = 1,5$; $SH_{rel}/600 = 0,98$).

Genetiivialkuosien hieman nominatiivia vähäisempi leksikaalinen monipuolisuus saattaa liittyä muodon taajaan käyttöön lajinnimityksissä (esim. *italianpapu*, *japaninsinappi*), joissa toistuvat usein samat maiden ja muiden alueiden nimet. Ryhmän yleisimmät alkuosat ovat kuitenkin *maailma* (30 kpl) ja *elämä* (25 kpl) (esim. *maailmanhalvaus* '[esim. eri maiden sisäpoliittisista syistä johtuva] kyvyttömyys ratkaista maailmanlaajuisia ongelmia valtioiden yhteisillä päätöksillä, *elämänrekisteri* 'älylaitteiden tietoverkkoihin tallentamat tiedot ihmisten elämästä'). Seuraavina tulevat *kansa* (17 kpl), *luonto* (16 kpl) ja *rauha* (12 kpl). Kaiken kaikkiaan ryhmässä on 865 yhdyssanaa, joissa esiintyy 521 eri alkuosaa.

Eri sijamuodoissa olevista alkuosista leksikaalisesti yksipuolisin on muiden kuin nominatiivien ja genetiivien ryhmä ($SLS = 1,5$, $H_{rel} = 0,9$). Muulla sijalla alkavia yhdyssanoja on kuitenkin aineistossa vain 28 kpl, joten niiden leksikaalisesta jakaumasta ei voida tehdä pitkälle meneviä päätelmiä. Ryhmän yleisimmät alkuosat ovat *maa* (6 kpl) ja *työ* (5 kpl).

Tässä luvussa olen tähän mennessä käsitellyt yhdyssanojen alkuosan leksikaalista monipuolisuutta alkuosan sanaluokan ja alkuosan muodon mukaan. Viimeisenä otan esille koko muodosteen sanaluokan, jolla on vielä alkuosan muotoa vähäisempi vaikutus alkuosien leksikaaliseen monipuolisuuteen. Taulukkoon 17 on koottu tiedot tutkimusaineiston yhdyssanojen alkuosien sane/lemma-suhteesta ($[S]SLS$) ja suhteellisesta entropiasta ($[S]H_{rel}$) koko muodosteen sanaluokan mukaan ryhmiteltynä. Lisäksi mukana ovat tiedot kaikkien ryhmään kuuluvien yhdyssanojen määrästä ja rakenteessa esiintyvien eri alkuosien määrästä sekä yleisimmistä alkuosalekseuseistä.

Kuten taulukosta 17 nähdään, eri sanaluokkiin kuuluvien yhdyssanojen alkuosan leksikaalisessa monipuolisuudessa ei juuri ole eroja ($SLS/100 = 1-1,2$, lisäksi taipumattomien ja vajaasti taipuvien sanaluokkien SLS 1,2; $SH_{rel}/100 = 0,99-1$, lisäksi taipumattomien ja vajaasti taipuvien sanaluokkien $H_{rel} = 0,97$). Koko muodosteen sanaluokka ei siis vaikuta yhdyssanan alkuosan leksikaaliseen monipuolisuuteen. Näin on siitä huolimatta, että eri sanaluokkiin kuuluvilla sanoilla on erilaiset yhdistymismahdollisuudet ja eri sanaluokkiin kuuluvien alkuosien leksikaalisessa monipuolisuudessa on melko selviä eroja (eri sanaluokkiin kuuluvien alkuosien leksikaalisesta monipuolisuudesta ks. tämän luvun alku; eri sanaluokkiin kuuluvien sanojen yhdistymisestä ks. luku 7.2.2).

TAULUKKO 17. Tutkimusaineiston yhdyssanojen alkuosan leksikaalinen monipuolisuus koko muodosteen sanaluokan mukaan

Sanaluokka	SSLS /100	SSLS /600	SLS	SH _{rel} /100	SH _{rel} /600	H _{rel}	Ryhmän yhdyssanat/ eri alkuosalekseemit	Yleisimmät alkuosalekseemit
Substantiivi	1	1,2	3,8	1	0,99	0,91	20 302/5 400	<i>euro</i> (156 kpl) <i>netti</i> (141 kpl)
Adjektiivi	1,1	1,4	1,7	0,99	0,96	0,95	1 035/625	<i>moni</i> (31 kpl) <i>uusi</i> (20 kpl)
Verbi	1,2	-	1,5	0,99	-	0,97	316/214	<i>pakko</i> (6 kpl) <i>syvä</i> (6 kpl) <i>täsmä</i> (6 kpl)
Partisiippi	1,1	-	1,2	0,99	-	0,98	245/202	<i>yli</i> (7 kpl) <i>etä-</i> (4 kpl) <i>itse</i> (4 kpl)
Useaan sanaluokkaan kuuluva (käytännössä substantiivi-adjektiivi)	1,1	-	1,2	0,99	-	0,98	177/148	<i>uusi</i> (8 kpl) <i>suomi</i> (4 kpl)
Taipumaton tai vajaasti taipuva sanaluokka (käytännössä adverbi)	-	-	1,2	-	-	0,97	68/58	<i>puoli</i> (6 kpl) <i>ennätys</i> (5 kpl)
Pronomini	-	-	1	-	-	1	1/1	<i>ei</i> (1 kpl)

6.2.3 Jälkiosa

Luvussa 6.2.1 totesin, että yhdyssanan jälkiosat ovat leksikaalisesti jonkin verran monipuolisempia kuin yhdyssanan alkuosat. Lisäksi yhdyssanan rakennepiirteillä on vähemmän vaikutusta jälki- kuin alkuosien leksikaaliseen monipuolisuuteen. Taulukoihin 18, 19 ja 20 on koottu tiedot yhdyssanan jälkiosan sane/lemma-suhteesta ([S]SLS) ja suhteellisesta entropiasta ($[S]H_{rel}$) alkuosan sanaluokan, alkuosan muodon ja koko muodosteen sanaluokan mukaan ryhmiteltynä. Lisäksi mukana ovat tiedot kaikkien ryhmään kuuluvien yhdyssanojen määrästä ja rakenteessa esiintyvien eri jälkiosien määrästä sekä yleisimmistä jälkiosaleksemeistä.

TAULUKKO 18. Tutkimusaineiston yhdyssanojen jälkiosan leksikaalinen monipuolisuus alkuosan sanaluokan mukaan

Alkuosan sanaluokka	SSLS /100	SSLS /600	SLS	SHrel /100	SHrel /600	Hrel	Ryhmän yhdyssanat/ eri jälkiosaleksemit	Yleisimmät jälkiosaleksemit
Substantiivi	1	1,1	3	1	0,99	0,93	17 680/5 812	<i>hoito</i> (72 kpl) <i>talous</i> (61 kpl)
Sanaluokaltaan epäspesifi	1	1,1	1,3	1	0,99	0,98	2 052/1 564	<i>kulttuuri</i> (10 kpl) <i>politiikka</i> (7 kpl) + muita 7 krt esiintyviä
Adjektiivi	1	1,1	1,2	1	0,99	0,99	1 238/1 034	<i>verkko</i> (6 kpl) <i>kieli</i> (5 kpl) <i>äiti</i> (5 kpl)
Useaa sanaluokkaa edustava	1	1,1	1,2	1	0,99	0,99	714/613	<i>vihreä</i> (5 kpl) <i>pop</i> (4 kpl) <i>työ</i> (4 kpl)
Taipumaton tai vajaasti taipuva sanaluokka	1	-	1,1	1	-	0,99	153/144	<i>otto</i> (4 kpl) <i>heitto</i> (2 kpl) + muita 2 krt esiintyviä
Verbi	1,1	-	1,1	0,99	-	0,99	108/98	<i>pala</i> (3 kpl) <i>kahva</i> (2 kpl) + muita 2 krt esiintyviä
Pronomini	1	-	1	1	-	1	101/98	<i>kieli</i> (3 kpl) <i>taide</i> (2 kpl)
Numeraali	-	-	1,1	-	-	1	90/85	<i>lappalainen</i> (2 kpl) <i>luokka</i> (2 kpl) + muita 2 krt esiintyviä
Partisiippi	-	-	1,1	-	-	0,98	8/7	<i>-silmäinen</i> (2 kpl) <i>ahneus</i> (1 kpl) + muita 1 krt esiintyviä

TAULUKKO 19. Tutkimusaineiston yhdyssanojen jälkiosan leksikaalinen monipuolisuus alkuosan muodon mukaan

Alkuosan muoto	SSLS /100	SSLS /600	SLS	SH _{rel} /100	SH _{rel} /600	H _{rel}	Ryhmän yhdyssanat / eri jälkiosalekseemit	Yleisimmät jälkiosalekseemit
Nominatiivi	1	1,1	3	1	0,99	0,93	17 599/5 857	<i>hoito</i> (70 kpl) <i>talous</i> (58 kpl)
Epäitsenäinen elementti	1	1,1	1,3	1	0,99	0,98	2 052/1 564	<i>kulttuuri</i> (10 kpl) <i>politiikka</i> (7 kpl) + muita 7 krt esiintyviä
Yhdysosa-muoto	1	1,1	1,3	1	0,99	0,99	1 359/1 081	<i>verkko</i> (7 kpl) <i>keskus</i> (6 kpl) + muita 6 krt esiintyviä
Genetiivi	1	1,2	1,3	1	0,98	0,98	865/664	<i>ruusu</i> (11 kpl) <i>hallinta</i> (10 kpl)
Taipumaton tai vaillinaisesti taipuva sana	1	-	1,1	1	-	0,99	241/225	<i>otto</i> (4 kpl) <i>euro</i> (3 kpl) <i>myynti</i> (3 kpl)
Muu sijamuoto	-	-	1,1	-	-	0,99	28/26	<i>muuttaja</i> (2 kpl) <i>otto</i> (2 kpl)

TAULUKKO 20. Tutkimusaineiston yhdyssanojen jälkiosan leksikaalinen monipuolisuus koko muodosteen sanaluokan mukaan

Sanaluokka	SSLS /100	SSLS /600	SLS	SH _{rel} /100	SH _{rel} /600	H _{rel}	Ryhmän yhdyssanat / eri jälkiosalekseemit	Yleisimmät jälkiosalekseemit
Substantiivi	1	1,2	3,4	1	0,99	0,93	20 302/6 010	<i>hoito</i> (78 kpl) <i>talous</i> (67 kpl)
Adjektiivi	1,1	1,5	1,8	0,99	0,96	0,95	1 035/577	<i>herkkä</i> (16 kpl) <i>-lähtöinen</i> (15 kpl)
Verbi	1,1	-	1,2	0,99	-	0,99	316/266	<i>kouluttaa</i> (5 kpl) <i>hoitaa</i> (4 kpl)
Partisiippi	1,1	-	1,3	0,99	-	0,98	245/187	<i>ohjattu</i> (6 kpl) <i>suuntautunut</i> (5 kpl) <i>värjätty</i> (5 kpl)
Useaan sanaluokkaan kuuluva (käytännössä substantiivi-adjektiivi)	1,6	-	1,9	0,96	-	0,94	177/94	<i>vihreä</i> (11 kpl) <i>rikas</i> (7 kpl) <i>-seksuaali</i> (7 kpl)
Taipumaton tai vajaasti taipuva sanaluokka (käytännössä adverbi)	-	-	1,1	-	-	0,99	68/62	<i>harvoin</i> (2 kpl) <i>mielessä</i> (2 kpl) + muita 2 krt esiintyviä
Pronomini	-	-	1	-	-	1	1/1	<i>mikä</i> (1 kpl)

Kuten taulukoista 18 ja 19 nähdään, alkuosan sanaluokalla ja alkuosan muodolla ei ole mainittavaa vaikutusta jälkiosien leksikaaliseen monipuolisuuteen. SLS- ja SH_{rel} -arvot ovat kaikissa ryhmissä samat tai lähes samat: alkuosan eri sanaluokissa $SLS/600 = 1,1$ ja $SLS/100 = 1-1,1$, $SH_{rel}/600 = 0,99$ ja $SH_{rel}/100 = 0,99-1$; alkuosan eri muotoryhmissä $SLS/600 = 1,1-1,2$ ja $SLS/100 = 1$, $SH_{rel}/600 = 0,98-0,99$ ja $SH_{rel}/100 = 1$. Myös koko muodosteen sanaluokan (taulukko 20) – joka on samalla myös jälkiosan oma sanaluokka – vaikutus on vähäinen ($SLS/600 = 1,2-1,5$ ja $SLS/100 = 1-1,6$; $SH_{rel}/600 = 0,96-0,99$ ja $SH_{rel}/100 = 0,96-1$). Kaiken kaikkiaan SLS- ja SH_{rel} -arvot osoittavat, että yhdyssanojen jälkiosat ovat leksikaalisesti varsin monipuolisia, eikä missään ryhmässä ole oleellisesti enempää variaatiota kuin toisessa. Edellisessä luvussa (6.2.2) esiin tulleet yhdyssanan alkuosan leksikaalista monimuotoisuutta koskevat rajoitukset eivät siis koske yhdyssanan jälkiosia.

Rahdun (1984: 410) arvion mukaan nominaalkuisille yhdysverbeille on tyyppillistä se, että niissä toistuvat samat yhdysosat. Tämän tutkimuksen aineistossa yhdysverbit eivät kuitenkaan eroa muihin sanaluokkiin kuuluvista yhdyssanoista sen enempää alku- (ks. taulukko 17 luvussa 6.2.2, s. 170) kuin jälkiosienkaan toistumisen kannalta (ks. taulukko 20 edellisellä sivulla). Myöskään pelkkien nominaalkuisten yhdysverbien SLS- ja H_{rel} -arvot eivät juuri eroa muista yhdysverbeistä. Pelkkien nominaalkuisten yhdysverbien ryhmässä yleisimmät jälkiosat ovat *kouluuttaa*, *lautailla* ja *maalata* (kutakin 3 kpl) ja alkuosat *pakko* ja *syvä* (kumpaakin 6 kpl). Yhdysverbien osien leksikaalisessa monimuotoisuudessa näyttää tapahtuneen muutoksia Rahdun (mt.) tutkimuksesta kuluneen kolmenkymmenen vuoden aikana. Muutokset voivat olla seurausta esimerkiksi kielenhuollon yhdysverbejä koskevien kantojen höllentymisestä.

6.2.4 Yhteenvedo

Uusien yhdyssanojen alkuosat ovat leksikaalisesti hieman yksipuolisempia kuin jälkiosat. Tämä on seurausta siitä, että yhdyssanan alkuosina käytetään jälkiosia enemmän niin sanottuihin muotosanojen sanaluokkiin kuuluvia sanoja, joita on kielessä niin sanottuja sisältösanoja vähemmän ja joissa on siten vähemmän myös leksikaalista monipuolisuutta.

Tässä tutkimuksessa tarkastelluilla yhdyssanan rakennepiirteillä on huomattavasti selvempi yhteys alku- kuin jälkiosien leksikaaliseen monipuolisuuteen. Sekä alku- että jälkiosien leksikaaliseen monipuolisuuteen vaikuttaa eniten osan itsensä sanaluokka. Jälkiosan kohdalla kuitenkin edes sen oman sanaluokan vaikutus ei ole järin suuri. Alkuosan sanaluokka ja muoto eivät juuri vaikuta jälkiosien leksikaaliseen monipuolisuuteen. Alkuosien leksikaaliseen monipuolisuuteen vaikuttaa niiden oman sanaluokan lisäksi jonkin verran myös niiden oma muoto. Tämä on odotuksenmukaista, sillä alkuosan muoto ja sanaluokka ovat jossain määrin riippuvaisia toisistaan. Koko muodosteen sanaluokalla (joka on samalla myös jälkiosan sanaluokka) on vain vähäinen vaikutus alkuosien leksikaaliseen monipuolisuuteen.

Tässä luvussa käytetyt tilastolliset tunnusluvut sane/lemma-suhde (SLS) ja suhteellinen entropia (H_{rel}) auttavat yhdessä tunnistamaan sellaiset rakenteet, joissa

rakenteessa esiintyy erityisen paljon tai vähän eri lekseemejä tai joissa eri lekseemit ovat erityisen tasaisesti tai epätasaisesti jakautuneita. Molemmilla mittareilla mitattuna yhdyssanan alkuosista leksikaalisesti monipuolisimpia ovat suomen uusille yhdyssanoille tyypillisimmät alkuosat eli nominatiivimuotoiset substantiivit.

Leksikaalista yksipuolisuutta voidaan tunnistaa kahdenlaista: a) sellaista, jossa ryhmässä esiintyy vain vähän eri lekseemejä ja b) sellaista, jossa eri lekseemit ovat jakautuneet epätasaisesti eli suuri osa rakenteen suosiosta on seurausta esimerkiksi yhden tai kahden frekventin alkuosan toistumisesta. Prefiksinkaltaiset alkuosaelementit ovat esimerkki ryhmästä, johon kuuluu verrattain vähän eri alkuosia mutta jossa eri alkuosat ovat kuitenkin jakautuneet melko tasaisesti. Selvimmin yhden sanan hallitsema on puolestaan pronominalkuosien ryhmä, jossa yli 70 % esiintymistä on *moni*-kvanttoripronomineja. Onkin paikallaan pohtia, onko uusissa yhdyssanoissa produktiiviseksi tulkittava enemmän *moni*-alkuinen rakenne kuin pronominalkuinen rakenne sinänsä. Muulla kuin *moni*-pronominilla alkavia sanoja on aineistossa vain 29.

Toinen ryhmä, jossa produktiivisuus voidaan katsoa enemmän tiettyjen alkuosalekseemien kuin rakenteen ominaisuudeksi, on muulla (kuin *-nen*-loppuisen sanan *-s*-loppuisella) konsonanttivartalolla alkavien yhdyssanojen joukko. Ryhmän yhteensä 151 yhdyssanassa esiintyy vain 7 eri alkuosaa. Vain hyvin harvat sanat voivat esiintyä yhdyssanan alussa muussa konsonanttivartalossa, eikä rakenteeseen saada myöskään uusia alkuosia.

Rakenteen leksikaalinen yksipuolisuus ei ole suora osoitus rakenteen epäproduktiivisuudesta eikä päätöstä rakenteen epäproduktiivisuudesta ei ole syytä tehdä ennen rakenteessa esiintyvien sanojen lähempää kvalitatiivista tarkastelua. Hyvä esimerkki tästä ovat *-inen*-loppuisten vierassanojen *-i*-loppuiset yhdysosamuodot, joita esiintyy yhteensä 165 yhdyssanan ryhmässä 30 erilaista. Näiden yhdysosamuotojen muodostus on mekaanista eivätkä *-inen*-loppuiset vierassanat esiinny suomen kielen uusien yhdyssanojen alkuosana muissa muodoissa. Rakenteeseen sopivia lekseemejä on kuitenkin ainakin toistaiseksi niin vähän, että variaatio jää vähäiseksi. Alkuosamuodon mekaanisen käytön ja uusista lainasanoista muodostettujen uusien *-inen*-loppuisten johdosten myötä myös rakenteen käyttö yhdyssananmuodostuksessa tulee todennäköisesti lisääntymään.

7 Rakennepiirteiden keskinäiset suhteet

Kahdessa edellisessä luvussa (5 ja 6) olen käsitellyt yksittäisten yhdyssanan rakennepiirteiden produktiivisuutta. Tässä luvussa siirryn tarkastelemaan piirteiden keskinäisiä vuorovaikutussuhteita eli sitä, mitkä piirteet esiintyvät todennäköisimmin yhdessä ja voiko tietyn rakennepiirteen käyttö edistää toisen rakennepiirteen valintaa tai jopa automaattisesti edellyttää sitä. Vuorovaikutussuhteita tarkastelemalla saadaan selville myös se, minkälaiset rakennepiirteiden yhdistelmät ovat suomen kielessä produktiivisessa käytössä.

Rakennepiirteiden keskinäisten suhteiden tarkastelu jakautuu kahteen osaan. Ensimmäisessä alaluvussa (7.1) käsitelen niitä yhdyssanarakenteen ominaisuuksia, jotka on koodattu tutkimusaineistoni kaikkiin yhdyssanoihin niiden leksikaalisten osien määrästä riippumatta. Toisessa alaluvussa (7.2) otan mukaan tarkasteluun myös ne ominaisuudet, jotka on koodattu vain kaksi leksikaalista osaa sisältäviin yhdyssanoihin. (Aineistoon koodatusta muuttujista tarkemmin luvussa 4.3.)

7.1 Yhdyssanan ja sen osien pituus

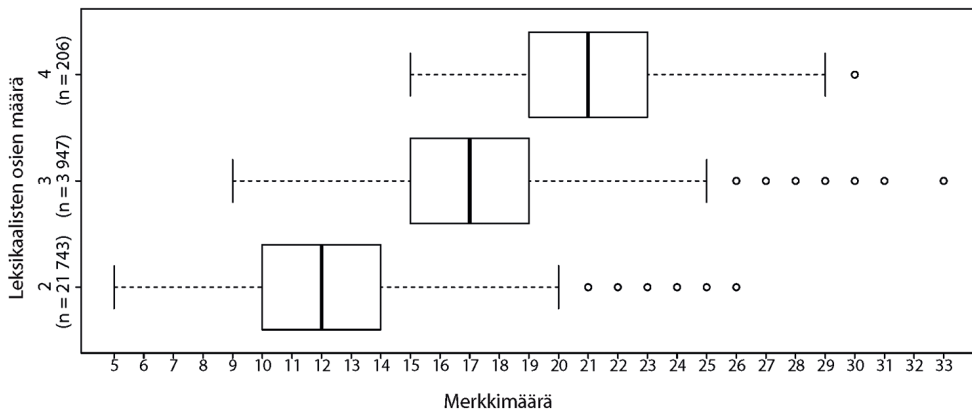
Tässä luvussa käsitelen niiden yhdyssanan rakenteellisten ominaisuuksien keskinäisiä vaikutussuhteita, jotka on koodattu tutkimusaineistoni kaikkiin yhdyssanoihin niiden leksikaalisten osien määrästä riippumatta. Nämä ominaisuudet ovat koko muodosteen sanaluokka, leksikaalisten osien määrä, koko muodosteen merkkimäärä sekä muodosteen ensimmäisen ja viimeisen osan merkkimäärä. Ensimmäisessä alaluvussa (7.1.1) tarkastelen yhdyssanan leksikaalisten osien määrän ja sekä koko muodosteen ja sen osien merkkimäärän suhdetta. Toisessa alaluvussa (7.1.2) keskityn uusien yhdyssanojen sanaluokan ja pituuden keskinäisiin vaikutussuhteisiin. Käytännössä tämä tarkoittaa sen selvittämistä, ovatko eri sanaluokkiin kuuluvat yhdyssanat tai niiden osat eripituisia.

Molemmissa alaluvuissa käytetyn aineiston pohjana on osa-aineisto 2, johon kuuluvat kaikki aineiston yhdyssanat lukuun ottamatta niitä, jotka sisältävät välimerkin tai lyhenteen (osa-aineistot on esitelty tarkemmin luvussa 4.3). 5 leksikaalista

osaa sisältävät yhdyssanat on jätetty tässä luvussa käytettävän osa-aineiston ulkopuolelle, sillä niitä on aineistossa vain 7 kpl (jäljelle jäävän aineiston $n = 25\ 896$). Lisäksi alaluvussa 7.2 aineistossa ovat mukana vain fennistiikassa perinteisesti käytettyihin sanaluokkiin kuuluvat sanat eli yhdyssubstantiivit, -adjektiivit, -verbit ja -adverbit (aineistossa ei ole muita taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvia sanoja kuin adverbeja, ks. tarkemmin luku 5.2.2). Myös aineiston ainoa yhdyspronomin on jätetty luvussa 7.2 käytettävän aineiston ulkopuolelle (jäljelle jäävän osa-aineiston $n = 25\ 446$).

7.1.1 Leksikaalisten osien määrä ja pituus

Jo edellä on todettu, että uusissa yhdyssanoissa on harvoin enemmän kuin 4 osaa ja että suomen kielen uusien yhdyssanojen mediaanipituus on 13 merkkiä (ks. luvut 5.1.1 ja 5.1.2). Yhdyssanan merkkimäärän ja leksikaalisten osien määrän suhdetta on havainnollistettu kuviossa 7. Kuviossa keskimmäinen paksu viiva osoittaa mediaanin paikan. Laatikon vasen reuna on alakvartiiliin ja oikea reuna yläkvartiiliin kohdalla. Laatikon päissä olevat janat ovat pituudeltaan $1,5 \times (Q_3 - Q_1)$, jossa Q_3 = yläkvartiili ja Q_1 = alakvartiili. Janan ulkopuolella olevat ympyrät ovat poikkeavia havaintoja. Tämä kuvaus pätee myös kaikkiin muihin tässä tutkimuksessa myöhemmin esiteltäviin laatikkojanakuvioihin.



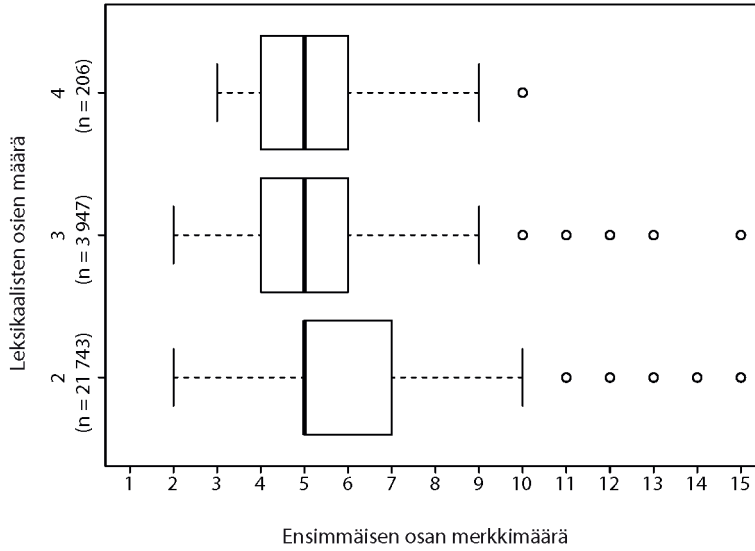
KUVIO 7. Leksikaalisten osien määrän ja koko muodosteen merkkimäärän suhde tutkimusaineiston yhdyssanoissa ($n = 25\ 896$)

Kuten kuviosta 7 nähdään, 2-osaisten yhdyssanojen merkkimäärän mediaani on 12 merkkiä, 3-osaisten 17 merkkiä ja 4-osaisten 21 merkkiä. Kuvion ja mediaanien perusteella vaikuttaa siis siltä, että yhdyssanan osat ovat sitä lyhyempiä, mitä pidempiä kokonaiset muodosteet ovat. Tämän puolesta puhuvat myös keskiarvot: 2-osaisten yhdyssanojen merkkimäärän keskiarvo on 12,4, jolloin yhden osan

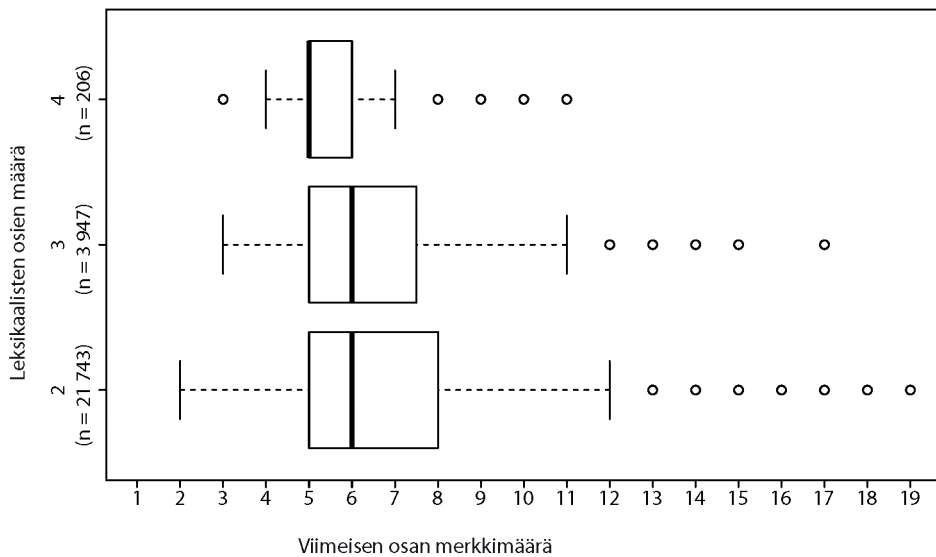
keskimääräinen pituus on 6,2 merkkiä. 3-osaisten merkkimäärän keskiarvo on 16,8 ja 4-osaisten 21,2 merkkiä, jolloin yhden osan keskimääräiseksi pituudeksi jää 3-osaisilla 5,6 ja 4-osaisilla vain 5,3 merkkiä. Kuviosta 8 ilmenee myös, että 2-, 3- ja 4-osaisten yhdyssanojen merkkimäärien hajonta on samantyyppistä (keskihajonta on 2-osaisilla 2,8 merkkiä, 3-osaisilla ja 4-osaisilla 2,9 merkkiä). Varianssien eroa testaavan Brownin–Forsythen testin mukaan ryhmien hajonnoissa ei ole tilastollisesti merkitsevää eroa ($n = 2206^{89}$, $F = 1,21$, $p = 0,4$). 2- ja 3-osaisten ryhmissä on kuitenkin enemmän pitkiä poikkeavia havaintoja kuin 4-osaisten ryhmässä.

Tarkastelen yhdyssanan osien merkkimäärän ja leksikaalisten osien määrän suhdetta vielä tarkemmin yksittäisten osien kautta. Käsittelen ensin yhdyssanan ensimmäistä leksikaalista osaa ja sen jälkeen viimeistä leksikaalista osaa. Vertailun helpottamiseksi uusien yhdyssanojen ensimmäisen ja viimeisen osan merkkimäärän ja leksikaalisten osien määrän suhdetta havainnollistavat kuviot on aseteltu allekkain samalle sivulle (kuviot 8 ja 9).

89 Testauksessa on tässä luvussa käytetty otoksia, joissa on kaikki aineiston 4-osaiset yhdyssanat (206 kpl) sekä 1000 satunnaisesti valittua 2- ja 3-osaista yhdyssanaa (tasakiintiöinen ositettu satunnaisotanta). Raportoidut tulokset ovat 10 otoksen keskiarvoja.



KUVIO 8. Leksikaalisten osien määrän ja ensimmäisen osan merkkimäärän suhde tutkimusaineiston yhdyssanoissa ($n = 25\ 896$)



KUVIO 9. Leksikaalisten osien määrän ja viimeisen osan merkkimäärän suhde tutkimusaineiston yhdyssanoissa ($n = 25\ 896$)

Kuviosta 8 (sivulla 178 yläpuolella oleva laatikkojanakuvi) nähdään, että sekä 2-, 3- että 4-osaisten yhdyssanojen ensimmäisen osan merkkimäärän mediaani on 5. Pelkän mediaanin⁹⁰ perusteella näyttää siis siltä, että ensimmäisten osien merkkimäärä ei muutu osien määrän kasvaessa. Keskiarvot näyttävät kuitenkin toista: 2-osaisten yhdyssanojen ensimmäisen osan merkkimäärän keskiarvo on 5,8, 3-osaisten 5,2 ja 4-osaisten 5. Keskiarvojen eron tilastollista merkitsevyyttä testaavan yksisuuntaisen testin⁹¹ mukaan tarkasteltavien ryhmien välillä on tilastollisesti merkitseviä eroja ($n = 2\ 206$, $df = 2$, $611,8$, $F = 41,81$, $p < 0,001^{***}$). Kahden riippumattoman otoksen t-testien avulla tehdyn parittaisen jatkotarkastelun mukaan 2-osaisten yhdyssanojen ensimmäisen osan merkkimäärä eroaa tilastollisesti merkitsevästi sekä 3- että 4-osaisten ensimmäisen osan merkkimäärästä (molempien vertailujen Bonferronin–Holmin korjauksella tarkistettu $p < 0,001^{***92}$). Sen sijaan 3- ja 4-osaisten ensimmäisen osan merkkimäärissä ei ole tilastollisesti merkitsevää eroa (tarkistettu $p = 0,2$).

Kuviosta 8 nähdään myös, että ensimmäisen osan merkkimäärän hajonta on sitä suurempi, mitä vähemmän leksikaalisia osia sanoissa on (2-osaisten keskihajonta on 1,8, 3-osaisten 1,6 ja 4-osaisten 1,4 merkkiä) ja että 2-osaisten ensimmäisen osan merkkimääräjakauma on selvästi vinompi kuin 3- ja 4-osaisten jakauma. 2-osaisten jakaumassa mediaani ja alakvartiili lankeavat yhteen. Yhtäaikaisesti usean ryhmän varianssien eroa testaavan Brownin–Forsythen testin mukaan ryhmien hajonnoissa on tilastollisesti merkitseviä eroja ($n = 2\ 206$, $F = 15,3$, $p < 0,001^{***}$). Parittaisen Brownin–Forsythen testien avulla tehdyn jatkotarkastelun mukaan 2-osaisten yhdyssanojen ensimmäisten osien merkkimäärän hajonta eroaa tilastollisesti merkitsevästi 3- ja 4-osaista (Bonferronin–Holmin korjauksella tarkistettu $p < 0,003^{**}$ verrattuna 3-osaisiin ja $p = 0,004^{**}$ verrattuna 4-osaisiin) mutta 3- ja 4-osaisten hajonnoissa ei ole keskenään tilastollisesti merkitsevää eroa (tarkistettu $p = 0,6$).

Tilastollisesti merkitsevät erot ensimmäisen osan merkkimäärän keskiarvoissa ja hajonnoissa ovat seurausta ensisijaisesti siitä, että 2-osaisten yhdyssanojen oikealle vinon jakauman pitkä häntä (ks. kuvio 9) ulottuu yhtenäisenä ja paksumpana pidemmälle kuin 3- ja 4-osaisten yhdyssanojen niin ikään oikealle vinot jakaumat. Lisäksi 2-osaissa yhdyssanoissa on suhteellisesti vähemmän hyvin lyhyitä (2–3 merkkiä pitkiä) ensimmäisiä osia kuin 3- ja 4-osaissa yhdyssanoissa.

Yhdyssanan viimeisen osan merkkimäärän ja koko muodosteen leksikaalisten osien määrän suhdetta havainnollistavasta kuviosta 9 (sivulla 178 alapuolella oleva laatikkojanakuvi) ilmenee, että viimeisten osien merkkimäärän mediaani on 2- ja 3-osaissa 6 merkkiä ja 4-osaissa 5 merkkiä. Viimeiset osat näyttävät eripituisilta myös keskiarvojen perusteella: 2-osaisten viimeisen osan merkkimäärän ka. = 6,6,

90 Koska jakaumat ovat hyvin vinoja, mediaani on jakaumien tarkastelussa periaatteessa keskiarvoa parempi tunnusluku. Aineiston suuren koon ansioista kuitenkin myös keskiarvo on käyttökelpoinen.

91 Testauksessa on käytetty R:n *oneway.test*-funktioita. Ks. tilastollisten menetelmien tarkempi esittely luvusta 4.4.

92 Parittaiset vertailut on tehty R:n *pairwise.t.test*-funktioilla, jonka tulostukseen ei kuulu t-testisuureen arvoja.

3-osaisten = 6,3 ja 4-osaisten = 5,7. Keskiarvojen eron tilastollista merkitsevyyttä testaavan yksisuuntaisen testin mukaan ryhmien välillä on tilastollisesti merkitseviä eroja ($n = 2\ 206$, $df = 2$, $650,3$, $F = 27,72$, $p < 0,001^{***}$). Viimeisten osien merkkimäärän parittaiset vertailut antavat samantyyppisen tuloksen kuin ensimmäisten osien, tosin sillä erotuksella, että siinä erot ovat tilastollisesti merkitseviä kaikkien kolmen ryhmän välillä (2- ja 3-osaisten välillä kahden riippumattoman otoksen t-testin Bonferronin–Holmin korjauksella tarkistettu $p = 0,03^*$; 2- ja 4-osaisten välillä $p < 0,001^{***}$; 3- ja 4-osaisten välillä $p < 0,001^{***}$).

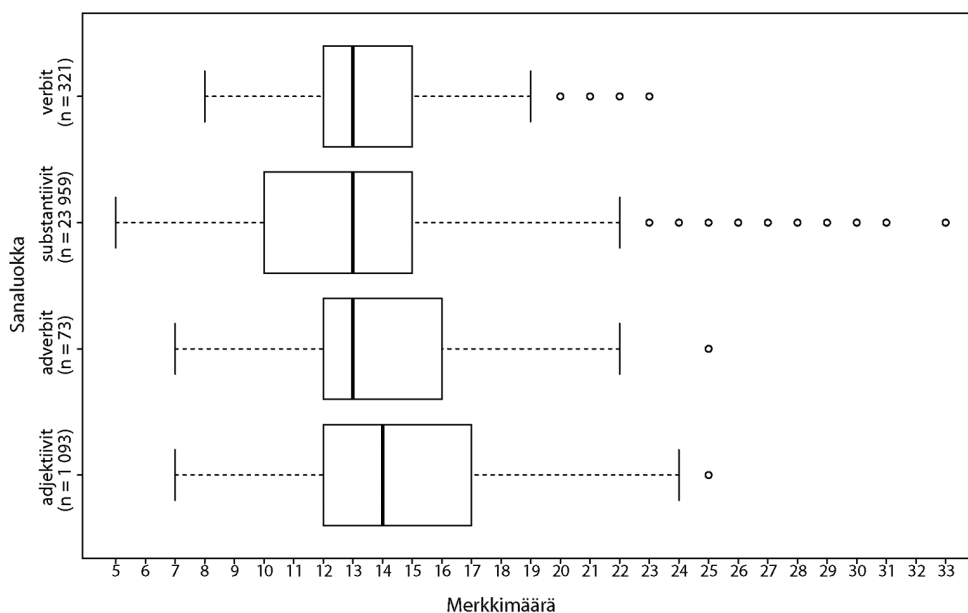
Kuviosta 9 nähdään myös, että 2- ja 3-osaisten viimeisen osan merkkimäärän jakaumat ovat keskenään samantyyppisiä. Kuitenkin myös viimeisten osien kohdalla 2-osaisten merkkimäärän hajonta on jonkin verran suurempi kuin 3-osaisten (2-osaisten viimeisen osan merkkimäärän keskihajonta on 2,2 ja 3-osaisten 2 merkkiä). 4-osaisten viimeisen osan merkkimääräjakauman muoto ja hajonta poikkeaa selvästi 2- ja 3-osaisten (4-osaisten merkkimäärän keskihajonta 1,5; mediaani ja alakvartiili lankeavat jakaumassa yhteen). Ensimmäisten osien tapaan myös viimeisten osien merkkimäärän hajonta on siis sitä suurempi, mitä vähemmän sanassa on osia. Yhtäaikaisesti usean ryhmän varianssien eroa testaavan Brownin–Forsythen testin mukaan ryhmien hajonnoissa on tilastollisesti merkitseviä eroja ($n = 2\ 206$, $F = 14,1$, $p < 0,001^{***}$). Parittaisten Brownin–Forsythen testien avulla tehdyn jatkotarkastelun mukaan 2- ja 3-osaisten yhdyssanojen viimeisen osan merkkimäärän hajonnat eivät eroa toisistaan tilastollisesti merkitsevästi (Bonferronin–Holmin korjauksella tarkistettu $p = 0,3$). Sen sijaan sekä 2- että 3-osaisten yhdyssanojen viimeisen osan merkkimäärän hajonta eroaa tilastollisesti merkitsevästi 4-osaisten yhdyssanojen viimeisen osan merkkimäärän hajonnasta (molempien vertailujen tarkistettu $p < 0,001^{***}$). Myös viimeisten osien kohdalla erot keskiarvossa ja hajonnassa ovat seurausta ensisijaisesti siitä, että oikealle vinojen jakaumien pitkät hännät (ks. kuvio 9) ulottuvat yhtenäisinä ja paksumpina sitä pidemmälle, mitä vähemmän osia on.

Tässä luvussa esitellyt tilastolliset analyysit osoittavat, että uusien yhdyssanojen osat ovat keskimäärin sitä lyhyempiä, mitä enemmän sanassa on osia⁹³. Hajontojen erot ryhmien välillä osoittavat, että vähän leksikaalisia osia sisältäviin yhdyssanoihin voidaan valita vapaammin eripituisia osia. Paljon osia sisältävissä sanoissa osien pituudessa on vähemmän vaihtelua esimerkiksi siksi, että niistä tulee helposti pitkiä ja hankalia käyttäjä, jos niissä on kovin pitkiä osia (ks. myös luku 2.3.3).

93 Mahdollisesti myös sanojen tavarakenteella ja puheessa lisäksi prosodisilla seikoilla voi olla vaikutusta siihen, millaisia osia eripituisiin yhdyssanoihin valitaan. Koska tämä tutkimus tarkastelee kirjoitettua kieltä, prosodisia seikkoja ei ole mahdollista tai tarpeellista ottaa huomioon. Myös yhdyssanan osien tavarakenne on rajattu tarkastelukohteiden ulkopuolelle.

7.1.2 Sanaluokka ja pituus

Jo edellä (luku 5.1.1) on todettu, että uusissa yhdyssanoissa on harvoin enemmän kuin 4 osaa ja että uudet yhdyssubstantiivit ovat tilastollisesti merkitsevästi lyhyempiä kuin Multasen (1991) tutkimat vuoden 1989 tilapäiset yhdyssubstantiivit. Lisäksi luvussa 5.1.2 on todettu, että suomen kielen uusien yhdyssanojen mediaanipituus on 13 merkkiä. Yhdyssanojen merkkimäärän ja sanaluokan suhdetta on havainnollistettu kuviossa 10.



KUVIO 10. Eri sanaluokkiin kuuluvien yhdyssanojen merkkimäärä tutkimusaineistossa ($n = 25\,446$)

Kuten kuviosta 10 ilmenee, mediaanilla mitattuna pisimpiä ovat yhdysadjektiivit, joiden mediaanipituus on 14 merkkiä. Sekä yhdyssubstantiivien, -adverbien että -verbien merkkimäärän mediaani on 13. Myös keskiarvolla mitattuna pisimpiä ovat adjektiivit (ka. = 14,6 merkkiä). Seuraavina tulevat adverbit (ka. = 14,2 merkkiä). Lyhimpiä ovat verbit (ka. = 13,2 merkkiä) ja substantiivit (ka. = 13 merkkiä). Keskiarvojen eron tilastollista merkitsevyyttä testaavan yksisuuntaisen testin⁹⁴ mukaan eri sanaluokkien keskimääräisessä merkkimäärässä on tilastollisesti merkitseviä

⁹⁴ Testauksessa on käytetty R:n *oneway.test*-funktioita. Ks. tilastollisten menetelmien tarkempi esittely luvusta 4.4.

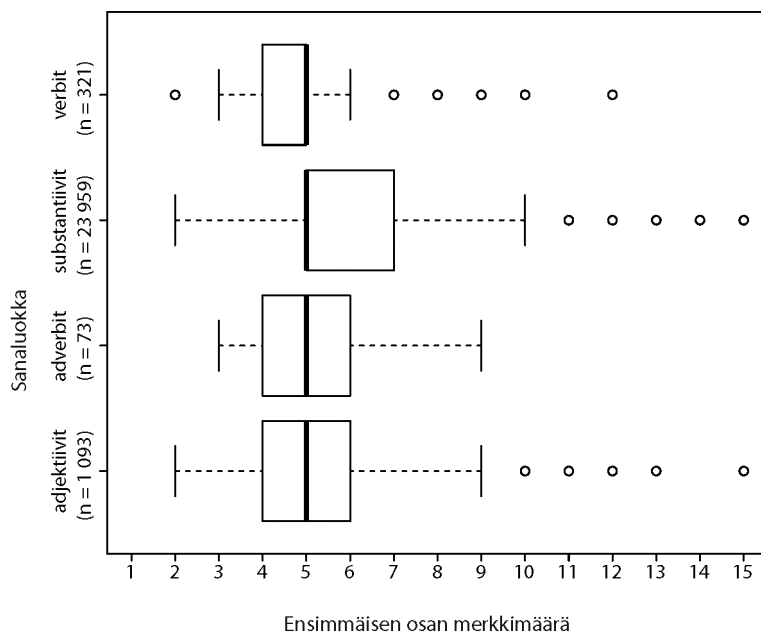
eroja ($n^{95} = 1\ 036$, $df = 3$, $294,1$, $F = 20,15$, $p < 0,001^{***}$). Kahden riippumattoman otoksen t-testien avulla tehdyn parittaisen jatkotarkastelun mukaan yhdysadjektiivit ovat tilastollisesti merkitsevästi pidempiä kuin yhdyssubstantiivit ja yhdysverbit (adjektiivien ja substantiivien välisen vertailun Bonferronin–Holmin korjauksella tarkistettu $p < 0,001^{***}$; adjektiivien ja verbien välisen vertailun tarkistettu $p < 0,001^{***96}$). Myös yhdysadverbit ovat tilastollisesti merkitsevästi pidempiä kuin yhdyssubstantiivit (tarkistettu $p = 0,016^*$). Muiden sanaluokkien välillä ei ole tilastollisesti merkitseviä eroja (adverbien ja verbien välisen vertailun tarkistettu $p = 0,07$; substantiivien ja verbien välisen vertailun tarkistettu $p = 0,37$; adjektiivien ja adverbien välisen vertailun tarkistettu $p = 0,4$).

Kuviosta 10 nähdään myös, että sekä pisimmät että lyhimmat tapaukset ovat aineistossa valtaenemmistönä olevia substantiiveja (keskihajonta 3,3 merkkiä). Myös adjektiivien keskihajonta on 3,3 merkkiä. Suurin keskihajonta on kuitenkin adverbeilla (3,4 merkkiä). Vähiten hajontaa on verbien pituudessa (keskihajonta 2,4 merkkiä). Yhtäaikaisesti usean ryhmän varianssien eroa testaavan Brownin–Forsythen testin mukaan ryhmien hajonnoissa on tilastollisesti merkitseviä eroja ($n = 1\ 036$, $F = 11,1$, $p < 0,001^{***}$). Parittaisten Brownin–Forsythen testien avulla tehty jatkotarkastelu osoittaa, että hajontaa koskevat erot ovat tilastollisesti merkitseviä vain verbien ja kaikkien muiden sanaluokkien välillä (verbien vertailussa sekä adjektiiveihin, adverbeihin että substantiiveihin Bonferronin–Holmin korjauksella tarkistettu $p < 0,001^{***}$, muiden vertailujen tarkistettu $p > 0,05$). Adjektiivien, adverbien ja substantiivien hajonnat eivät siis poikkea tilastollisesti merkitsevästi toisistaan.

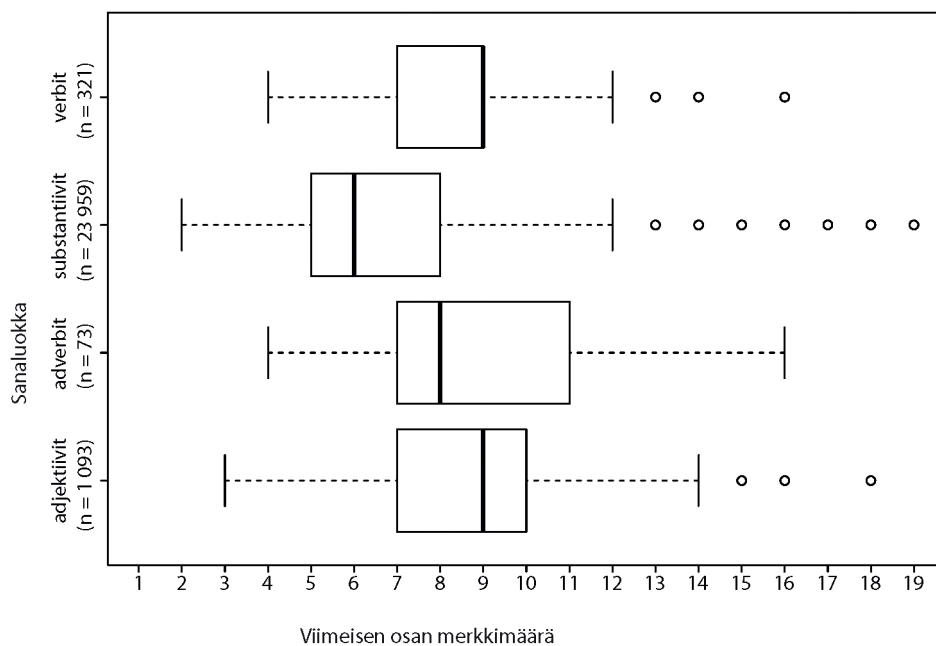
Sanaluokkien välillä olevien erojen syyt paljastuvat parhaiten tarkasteltaessa muodosteiden ensimmäisten ja viimeisten osien pituutta. Käsittelen ensin yhdyssanan ensimmäistä leksikaalista osaa ja sen jälkeen viimeistä leksikaalista osaa. Vertailun helpottamiseksi uusien yhdyssanojen ensimmäisen ja viimeisen osan merkimmäärän ja sanaluokan suhdetta havainnollistavat kuviot on aseteltu allekkain samalle sivulle (kuviot 11 ja 12).

95 Tämän luvun testauksessa on käytetty otoksia, joissa on kaikki aineiston yhdysadverbit (73 kpl) ja kaikki aineiston yhdysverbit (321 kpl) sekä 321 satunnaisesti valittua yhdysadjektiivia ja -substantiivia (tasakiintiöinen ositettu satunnaisotanta) ($73 + 3 \times 321 = 1036$). Raportoidut tulokset ovat 10 otoksen keskiarvoja.

96 Parittaiset vertailut on tehty R:n *pairwise.t.test*-funktioilla, jonka tulostukseen ei kuulu t-testisuurteen arvoja.



KUVIO 11. Eri sanaluokkiin kuuluvien yhdyssanojen ensimmäisen osan merkkimäärä tutkimusaineistossa ($n = 25\,446$)



KUVIO 12. Eri sanaluokkiin kuuluvien yhdyssanojen viimeisen osan merkkimäärä tutkimusaineistossa ($n = 25\,446$)

Kuten kuvio 11 (sivulla 183 yläpuolella oleva laatikkojanakuvio) osoittaa, yhdyssanojen ensimmäisen osan mediaanipituus on sanaluokasta riippumatta 5 merkkiä. Yhdyssubstantiivien jakaumassa mediaani ja alakvartiili lankeavat yhteen. Yhdysverbien jakaumassa puolestaan mediaani ja yläkvartiili lankeavat yhteen. Keskiarvon perusteella pisimmät alkuosat löytyvät substantiiveista (ka. = 5,7), vaikka yhdyssubstantiivit ovat koko muodosteen tasolla tarkasteltuna eri sanaluokkiin kuuluvista yhdyssanoista lyhimpiä. Tämän poikkeuksen jälkeen ensimmäisten osien pituusjärjestys on sama kuin kokonaisten muodosteidenkin: adjektiivien ensimmäisten osien merkkimäärän ka. = 5,4, adverbien ka. = 5 ja verbien ka. 4,8.

Keskiarvojen eron tilastollista merkitsevyyttä testaavan yksisuuntaisen testin mukaan eri sanaluokkiin kuuluvien yhdyssanojen ensimmäisen osan keskimääräisessä merkkimäärässä on tilastollisesti merkitseviä eroja ($n = 1\ 036$, $df = 3$, 313,4, $F = 22,3$, $p < 0,001^{***}$). Kahden riippumattoman otoksen t-testien avulla tehdyn parittaisen jatkotarkastelun mukaan yhdyssubstantiivien ensimmäiset osat ovat tilastollisesti merkitsevästi pidempiä kuin yhdysverbien ja -adverbien (substantiivien ja verbien välisen vertailun Bonferronin–Holmin korjauksella tarkistettu $p < 0,001^{***}$; substantiivien ja adverbien välisen vertailun tarkistettu $p = 0,022^*$). Myös yhdysadjektiivien ensimmäiset osat ovat tilastollisesti merkitsevästi pidempiä kuin yhdysverbien alkuosat (tarkistettu $p < 0,001^{***}$). Muiden sanaluokkien välillä ei ole tilastollisesti merkitseviä eroja (adjektiivien ja adverbien välisen vertailun tarkistettu $p = 0,09$; adverbien ja verbien välisen vertailun tarkistettu $p = 0,32$; substantiivien ja adjektiivien välisen vertailun tarkistettu $p = 0,37$).

Eröt eri sanaluokkiin kuuluvien yhdyssanojen ensimmäisten osien merkkimäärässä eivät selitä sitä edellä mainittua seikkaa, että yhdyssubstantiivit ovat kokonaisina muodosteinä keskimäärin lyhyempiä kuin yhdysadjektiivit ja -adverbit. Sen sijaan yhdysverbien lyhyttä selittävät niiden lyhyet alkuosat. Jopa 18,2 % yhdysverbien alkuosista on erilaisia 3–4 merkkiä pitkiä kotoisia ja vierasalkuisia prefiksinkaltaisia osia (esim. *pikakäännyttää*, *ylinaisittaa*, *biomuuntua*).

Kuviosta 11 nähdään myös, että kaikkien sanaluokkien ensimmäisen osan merkkimääräjakaumat ovat oikealle vinoja. Ensimmäisen osan pituusjakaumien sijainnissa ja hajonnassa on kuitenkin sanaluokkakohtaisia eroja. Suurin keskihajonta on yhdyssubstantiiveilla ja -adjektiiveilla (molemmilla 1,8 merkkiä) ja pienin yhdysadverbeilla ja -verbeilla (molemmilla 1,3 merkkiä). Ainakaan ensimmäiset osat eivät näin ollen selitä sitä, että kokonaisissa muodosteissa adverbien merkkimäärän hajonta on suurinta (ks. edellä). Yhtäaikaisesti usean ryhmän varianssien eroa testaavan Brownin–Forsythen testin mukaan ryhmien hajonnoissa on tilastollisesti merkitseviä eroja ($n = 1\ 036$, $F = 6,7$, $p < 0,001^{***}$). Parittaisten Brownin–Forsythen testien avulla tehty jatkotarkastelu osoittaa erojen olevan tilastollisesti merkitseviä substantiivien ja verbien (Bonferronin–Holmin korjauksella tarkistettu $p < 0,001^{***}$) sekä adjektiivien ja verbien (tarkistettu $p = 0,006^{**}$) välillä. Muiden sanaluokkien välillä eivät ensimmäisen osan merkkimäärän hajonnassa olevat erot ole tilastollisesti merkitseviä (kaikkien muiden parien tarkistettu $p > 0,05$).

Yhdyssubstantiivien ja -adjektiivien ensimmäisen osan merkkimäärän yhdysverbejä suurempi hajonta osoittaa, että näiden sanaluokkien alkuosiksi voidaan valita vapaammin eripituisia leksikaalisia osia. Tähän viittaa myös se, että yhdysubstantiivien ja -adjektiivien jakauman pitkä häntä (ks. kuvio 11) ulottuu pidemmälle oikealle kuin yhdysverbien ja -adverbien. Lisäksi yhdyssubstantiivien ja -adjektiivien ryhmissä on enemmän pitkiä poikkeavia havaintoja kuin yhdysverbien ja -adverbien ryhmissä.

Yhdyssanojen viimeisten osien tarkastelussa on olennaista muistaa, että koko yhdysosan sanaluokka on sama kuin sen jälkiosan sanaluokka (ks. luvut 5.2.2 ja 5.2.3). Yhdyssosan sanaluokan ja viimeisen osan merkkimäärän suhdetta havainnollistava kuvio 12 (sivulla 183 alapuolella oleva laatikkojanakuvio) osoittaa, että viimeisten osien keskimääräisessä merkkimäärässä on enemmän vaihtelua sanaluokkien välillä kuin ensimmäisten osien merkkimäärässä. Korkein mediaani on yhdysadjektiivien ja -verbien viimeisillä osilla (molemmilla 9 merkkiä). Seuraaviksi pisimpiä ovat yhdysadverbien (8 merkkiä) ja lyhimpiä yhdyssubstantiivien (6 merkkiä) viimeiset osat. Keskiarvolla mitattuna pisimmät viimeiset osat on yhdysadjektiiveissa ja yhdysverbien sijaan yhdysadverbeissa (sekä adjektiivien että adverbien ka. = 8,8). Yhdysverbien viimeisen osan merkkimäärän keskiarvo on 8,3 ja selvästi lyhimpien eli yhdyssubstantiivien 6,4.

Keskiarvojen eron tilastollista merkitsevyyttä testaavan yksisuuntaisen testin mukaan eri sanaluokkiin kuuluvien yhdyssanojen viimeisen osan keskimääräisessä merkkimäärässä on tilastollisesti merkitseviä eroja ($n = 1\ 036$, $df = 3$, $290,5$, $F = 86,8$, $p < 0,001^{***}$). Kahden riippumattoman otoksen t-testien avulla tehdyn parittaisen jatkotarkastelun mukaan yhdyssubstantiivien viimeiset osat ovat tilastollisesti merkitsevästi lyhyempiä kuin kaikkiin muihin sanaluokkiin kuuluvien yhdyssanojen (substantiivien ja kaikkien muiden sanaluokkien välisten vertailujen Bonferroinin–Holmin korjauksella tarkistettu $p < 0,001^{***}$). Yhdysadjektiivien viimeiset osat ovat myös yhdysverbien viimeisiä osia tilastollisesti merkitsevästi pidempiä (tarkistettu $p = 0,012^*$). Muiden sanaluokkien välillä ei ole tilastollisesti merkitseviä eroja (adjektiivien ja adverbien välisen vertailun tarkistettu $p = 0,78$; adverbien ja verbien välisen vertailun tarkistettu $p = 0,29$).

Yhdyssubstantiivien lyhyet viimeiset osat selittävät kokonaisten yhdyssubstantiivien keskimääräistä lyhemmyyttä. Vaikka aineiston pisin yhdysosan viimeinen osa on substantiivi (19-merkkinen jälkiosa sanassa *uusbyrokratisoituminen*), tyypilliset yhdyssubstantiivien viimeiset osat ovat lyhyitä johtamattomia perussanoja (esim. *talvikoulu*, *unisukka*). Pitkistä johdoksista tavallisimpia ovat *-minen*-johdokset, joita on noin 1,5 %:ssa aineiston yhdyssubstantiiveista (≥ 10 merkkiä pitkä *-minen*-johdos on n. 0,9 %:ssa kaikista yhdyssubstantiiveista).

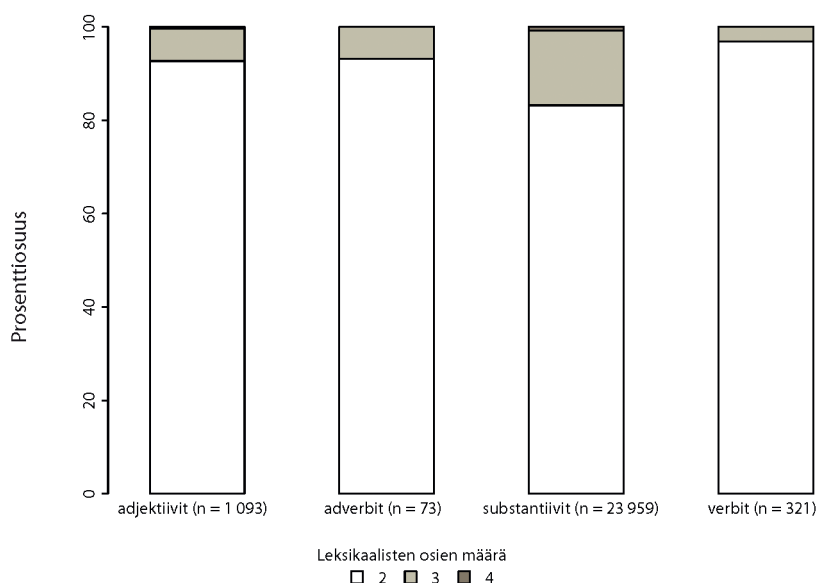
Kokonaisten yhdysadjektiivien keskimääräistä pidemmyyttä selittävät niiden pitkät viimeiset osat. Suomen kielen adjektiivit ovat useissa viestintälajeissa keskimäärin pidempiä kuin substantiivit, verbit ja adverbis (kaikkien sanaluokkien sanoista vain numeraalit ovat keskimäärin adjektiiveja pidempiä, Niemikorpi 1990: 88–89). Yhdysadjektiivien viimeisissä osissa esiintyy paljon sellaisia johtimia, jotka

ovat jo itsessään yhtä pitkiä kuin tyypilliset yhdyssanan leksikaaliset osat (vrt. yhdyssanan jälkiosan mediaanipituus 6 merkkiä ja esim. 6 merkkiä sisältävä *-mAinen-johdin*). Yleisiä ovat esimerkiksi *-llinen-* (9,4 % kaikista yhdysadjektiiveista) ja *-mAinen-johdokset* (1,7 % kaikista yhdysadjektiiveista) (esim. *käsialallinen*, *maise-makonttorimainen*). Jopa noin 75 % uusien yhdysadjektiivien viimeisistä osista on erilaisia *-inen-johdoksia* (n. 34 % yhdysadjektiiveista päättyy ≥ 10 merkkiä pitkään *-inen-johdokseen*).

Kuviosta 12 nähdään myös, että viimeisten osien merkkimäärät eivät ole normaalisti jakautuneita mutta jakaumat eivät ole yhtä selvästi oikealle vinoja kuin ensimmäisten osien merkkimääräjakaumat (vrt. kuvio 11). Suurin keskihajonta on adverbeilla (2,7 merkkiä). Seuraavina tulevat adjektiivit (2,4 merkkiä), substantiivit (2 merkkiä) ja verbit (1,8 merkkiä).⁹⁷ Kokonaisten adverbien suuri merkkimäärähajonta on siis seurausta ensisijaisesti niiden viimeisten osien suuresta hajonnasta (minimi 4 ja maksimi 16). Yhdysadverbien viimeisiin osiin kuuluu poikkeuksellisen paljon pitkiä ja lyhyitä sanoja (esim. *yhteistoiminnallisesti*, *lämpökontrolloidusti* ja *tihkipäättä*, *ennätysylös*; ts. niiden joukossa on vain vähän keskimittaisia sanoja). Pitkiä ovat varsinkin adjektiiveista, partisiipeista ja verbeistä johdetut *-sti-*adverbit, joita on kaikista yhdysadverbeista 40,3 % (n. 29 % yhdysadverbeista päättyy ≥ 10 merkkiä pitkään *-sti-*loppuiseen adverbiin).

Edellä tarkastellun leksikaalisten osien pituuden lisäksi olennainen kokonaisten yhdyssanojen pituutta selittävä seikka on tietysti niiden osien määrä. Uusien yhdysanojen sanaluokan ja osien määrän suhdetta on havainnollistettu kuviossa 13.

97 Yhtäaikaisesti usean ryhmän varianssien eroa testaavan Brownin–Forsythen testin mukaan ryhmien hajonnoissa on tilastollisesti merkitseviä eroja ($n = 1\ 036$, $F = 9,1$, $p < 0,001^{***}$). Parittaisten Brownin–Forsythen testien avulla tehty jatkotarkastelu osoittaa erojen olevan tilastollisesti merkitseviä kaikkien muiden sanaluokkien paitsi adjektiivien ja adverbien (Bonferronin–Holmin korjauksella tarkistettu $p = 0,4$) ja substantiivien ja verbien (tarkistettu $p = 0,3$) välillä. Sen sijaan adjektiivien ja substantiivien (tarkistettu $p = 0,008^{**}$), adjektiivien ja verbien (tarkistettu $p < 0,001^{***}$), adverbien ja substantiivien (tarkistettu $p = 0,014^*$) sekä adverbien ja verbien (tarkistettu $p < 0,001^{***}$) hajonnat eroavat toisistaan tilastollisesti merkitsevästi.



KUVIO 13. Eri sanaluokkiin kuuluvien yhdyssanojen leksikaalisten osien määrä tutkimusaineistossa ($n = 25\,446$)

Kuten kuviosta 13 nähdään, osien määrällä mitattuna pisimpiä yhdyssanoja ovat substantiivit. Substantiivien ryhmässä on suhteellisesti vähemmän 2-osaisia (83,2 %) ja enemmän 3- ja 4-osaisia (3-osaisia 16 % ja 4-osaisia 0,8 %) tapauksia kuin muissa sanaluokissa. Adjektiiveista 92,7 % on 2-osaisia, 7 % 3-osaisia ja 0,4 % 4-osaisia. Adverbeissa osuudet ovat samansuuntaiset sillä erotuksella, että niiden ryhmässä ei ole lainkaan 4-osaisia tapauksia (2-osaisia 93,2 % ja 3-osaisia 6,8 %). Suhteellisesti eniten 2-osaisia (96,9 %) ja vähiten 3-osaisia tapauksia (3,1 %) on verbien ryhmässä. Myöskään verbien ryhmässä ei ole lainkaan 4-osaisia tapauksia.

Fisherin tarkan testin mukaan eri sanaluokkiin kuuluvien yhdyssanojen leksikaalisten osien määrässä⁹⁸ on tilastollisesti merkitseviä eroja ($n = 1036$, $p < 0,001^{***}$). Sovitettujen standardoitujen jäännösten avulla tehdyn jatkotarkastelun mukaan substantiivien joukossa on tilastollisesti merkitsevästi enemmän 3-osaisia tapauksia kuin kaikissa yhdyssanoissa keskimäärin (sovitettu standardoitu jäännös = 4,6) ja verbien joukossa on tilastollisesti merkitsevästi vähemmän 3-osaisia tapauksia kuin kaikissa yhdyssanoissa keskimäärin (sovitettu standardoitu jäännös = -3,4). Muuten eri sanaluokkiin kuuluvien yhdyssanojen leksikaalisten osien määrä ei poikkea tilastollisesti merkitsevästi kaikkien sanaluokkien yhteisistä keskiarvoista (sovitettut standardoidut jäännökset > 2 tai < -2).

98 Leksikaalisten osien määrää on tässä käsitelty kategorisena muuttujana, jonka luokkia ovat 2-osaiset, 3-osaiset ja 4-osaiset yhdyssanat. Koska muuttujalla on vain kolme mahdollista arvoa, sen käsittely numeerisena muuttujana ei välttämättä antaisi luotettavia tuloksia.

Sen lisäksi, että yhdyssubstantiiveissa on keskimäärin enemmän osia kuin muihin sanaluokkiin kuuluvissa yhdyssanoissa, kaikki tämän tutkimuksen aineiston 5-osaiset yhdyssanat (koko tutkimusaineistossa 7 kpl; jätetty tässä luvussa käytetyn osa-aineiston ulkopuolelle) ovat substantiiveja. 4-osaisiin yhdyssanoihin kuuluu substantiivien lisäksi myös adjektiiveja. Sitä, pidetäänkö muihin sanaluokkiin kuuluvia paljon leksikaalisia osia sisältäviä yhdyssanoja ylipäänsä kielenmuokaisina, olisi tutkittava erikseen. Kaikkiin sanaluokkiin kuuluvien pitkien yhdyssanojen muodostamista rajoittaa todennäköisesti ensisijaisesti käytettävyyks (ks. myös luku 5.1.1). Käytettävyyden rajat näyttävät tulevan muihin sanaluokkiin kuuluvien yhdyssanojen kohdalla vastaan nopeammin kuin substantiivien kohdalla.

Tässä luvussa esitellyt eri sanaluokkiin kuuluvien sanojen osien määrää koskevat tilastolliset analyysit osoittavat, että yhdysadjektiivien merkkimäärällä mitattu pidemmys ei ole seurausta niiden osien määrästä. Ainoaksi selittäjäksi yhdysadjektiivien pituudelle jäävät siis niiden keskimääräistä pidemmät jälkiosat. Osien määrä ei selitä myöskään yhdyssubstantiivien lyhyttä: yhdyssubstantiivit ovat eri sanaluokkien yhdyssanoista keskimäärin lyhimpiä, vaikka niiden joukossa on suhteellisesti eniten enemmän kuin kaksi leksikaalista osaa sisältäviä tapauksia. Yhdysubstantiivien lyhyden ainoaksi selittäjäsi jää siis niiden jälkiosien lyhyys. Osien määrän sanaluokittainen tarkastelu sen sijaan osoittaa, että yhdysverbien lyhyys on niiden lyhyiden alkuosien lisäksi seurausta niiden leksikaalisten osien vähyydestä.

7.1.3 Yhteenveto

Yhdyssanan merkkimäärä, leksikaalisten osien määrä ja sanaluokka sekä muodosteen osien merkkimäärä ovat kaikki yhteydessä toisiinsa. Odotuksenmukaisesti yhdyssanan merkkimäärä kasvaa sen leksikaalisten osien määrän lisääntyessä. Lisäys ei kuitenkaan ole suoraviivaista, sillä yhdyssanan osat ovat sitä lyhyempiä, mitä enemmän niitä on. Myös osien merkkimäärän hajonta pienenee osien määrän lisääntyessä. Näyttääkin siltä, että yhdyssanaan voidaan valita eripituisia ja varsin pitkiäkin osia sitä vapaammin, mitä vähemmän sanassa on leksikaalisia osia. Vastaavasti mitä enemmän osia yhdyssanassa on, sitä lyhyempiä osien on oltava, jotta muodosteesta ei tule kokonaisuutena liian pitkää ja hankalaa käytettäväksi. Pitkänkin yhdyssanan hahmottaminen ja käyttö helpottuvat, jos muodosteen osana on jo valmiiksi leksikaalistuneita lyhyitä yhdyssanoja (esim. *elokuva*, *yhteiskunta*, *periaate* jne.). Tällaiset osat ovatkin yleisiä pitkissä yhdyssanoissa.

Eri sanaluokkiin kuuluvien yhdyssanojen merkkimäärällä mitatussa keskipituudessa on tilastollisesti merkitseviä eroja. Keskimäärin pisimpiä ovat adjektiivit. Toiseksi pisimpiä ovat adverbit, kolmanneksi pisimpiä verbit ja lyhimpiä substantiivit. Leksikaalisten osien määrällä mitattuna puolestaan pisimpiä ovat substantiivit. Niitä seuraavat adjektiivit ja adverbit ja lyhimpinä verbit. Syyt eri sanaluokkiin kuuluvien yhdyssanojen merkkimäärällä mitatulle pituudelle tai lyhyydelle ovatkin erilaisia: Adjektiivien pituutta selittävät niiden pitkät viimeiset osat mutta ei niiden leksikaalisten osien määrä, substantiivien lyhyttä niiden viimeisten osien lyhyys mutta ei osien määrä ja verbien lyhyttä sekä osien määrä että alkuosan lyhyys.

7.2 Selitettävänä alkuosan muoto ja sanaluokka

Tässä luvussa käsittelen tutkimusaineistoni 2-osaisiin yhdyssanoihin koodattujen muuttujien keskinäisiä suhteita. Tarkastelussa mukana olevat muuttujat ovat alkuosan muoto (karkea jaottelu, ks. luku 5.3.1), alkuosan sanaluokka (ks. luku 5.2.4), yhdyssanan merkkimäärä (ks. luku 5.1.2) ja koko muodosteen sanaluokka (ks. luku 5.2.2). Vuorovaikutussuhteiden tarkastelun kautta saadaan selville, millaiset rakennepiirteiden yhdistelmät ovat produktiivisia suomen kielen yhdyssanoissa ja mitkä piirteet esiintyvät uudessa yhdyssanastossa toisia todennäköisemmin yhdessä.

Yhdyssanojen rakennepiirteiden keskinäisiä suhteita kuvattaessa on tärkeää muistaa, että vaikutussuhde ei ole sama asia kuin syy–seuraussuhde. Sananmuodostuksessa kaikki ominaisuudet vaikuttavat toisiinsa, eikä eri ominaisuuksia voida siksi pitää yksiselitteisesti joko toistensa syinä tai seurauksina (ks. myös Lehecka 2012: 71). Tässä luvussa pääasiallisena menetelmänä käytetyissä luokittavissa puissa (tutkimuksessa käytettävistä tilastollisista menetelmistä ks. tarkemmin luku 4.4) yksi muuttujista on kuitenkin valittava selitettäväksi ja muut selittäviksi. Koska luokittavissa puissa selitettäviä muuttujia voi olla vain yksi kerrallaan, olen tehnyt analyysin kahdesta eri näkökulmasta. Tämä mahdollistaa muuttujien välisten suhteiden monipuolisemman tarkastelun.

Ensimmäisessä alaluvussa (7.2.1) olen valinnut selitettäväksi muuttujaksi alkuosan morfologisen muodon. Se on produktiivisuuden kannalta mielenkiintoisin muuttuja ainakin kahdesta syystä: 1) alkuosan muotoa on pidetty vanhastaan hyvin olennaisena suomen yhdyssanojen rakenteeseen liittyvänä kysymyksenä, joka on herättänyt paljon keskustelua esimerkiksi kielenhuollossa (esim. Kettunen 1949: 103–113; Vesikansa 1989: 227–237; T. Kolehmainen 2011, 2014: 286–299) ja 2) alkuosan muodon valinta kertoo paljon siitä, miten informaatiota halutaan tiivistää ja minkälaisen morfologisten vihjeiden katsotaan riittävän yhdyssanan oikeaan tulkintaan (ks. luvut 2.3.2 ja 5.3.1).

Toisessa alaluvussa (7.2.2) selitettävänä muuttujana on yhdyssanan alkuosan sanaluokka ja tarkastelun kohteena erityisesti sen yhteys koko muodosteen sanaluokkaan. Selitettävänä muuttujana olisi voinut yhtä hyvin olla myös koko muodosteen sanaluokka, mutta eri malleja kokeiltuani valitsin niistä helpommin hahmotettavan. Myös eri sanaluokkien yhdistymismahdollisuuksia on pidetty oleellisena suomen yhdyssanojen rakenteeseen liittyvänä kysymyksenä (esim. ISK § 407). Suomen kielopin sanotaan rajoittavan yhdyssanojen muodostusta vain vähän, ja eri sanaluokkiin kuuluvien sanojen yhdistymiselle on vain harvoja esteitä (esim. persoonamuotoisten verbien yhdistäminen summayhdyssanaksi ei ole mahdollista, vrt. joidenkin suomen sukukielten ns. parisanat, Erkki Itkonen 1966: 236–237). Kuitenkin – kuten luvussa 5.2.4 on jo todettu – eri sanaluokkiin kuuluvien sanojen yhdistymismahdollisuudet ja yhdistämisen vaikutus osien eri sanaluokkia edustavien sanojen muotoon ovat erilaisia, ja vaikka periaatteellisia rajoituksia on vähän, käytännössä suurin osa niin suomen kielen yhdyssanoista kuin niiden alkuosistakin on substantiiveja.

Luvussa käytetty aineisto pohjaa pelkästään kaksiosaisia yhdyssanoja sisältävään osa-aineistoon 3 (osa-aineistot on esitelty tarkemmin luvussa 4.3). Eri sanaluokkiin kuuluvista yhdyssanoista tässä luvussa käytetyssä aineistossa ovat mukana selvästi yhteen perinteisistä sanaluokista kuuluvat sanat (yhdyssubstantiivit, -adjektiivit, -verbit ja -adverbit; osa-aineiston ulkopuolelle on jätetty sanaluokaltaan häilyvät substantiivi-adjektiivit ja partisiipit sekä aineiston yksi pronomini). Alkuosan sanaluokista osa-aineiston ulkopuolelle on jätetty partisiipit, joita on aineistossa vain 8 kpl (jäljelle jäävän osa-aineiston $n = 21\ 713$).

7.2.1 Muiden rakennepiirteiden yhteys alkuosan muotoon

Tässä luvussa käsittelen yhdyssanan muiden rakennepiirteiden yhteyttä yhdyssanan alkuosan muotoon. Esittelen luvussa useita rakennepiirteiden välisiä keskinäisiä suhteita kuvaavia puumalleja. Ensimmäisessä puumallissa selittävinä muuttujina ovat alkuosan sanaluokka, koko muodosteen sanaluokka ja yhdyssanan pituus. Koska luokittavat puut kuvaavat havainnollisimmin mallin tärkeimmäksi nostaman selittävän muuttujan vaikutusta selitettävän muuttujan jakaumaan (luokittavista puista ks. tarkemmin luku 4.4), seuraavasta puusta on pudotettu pois edellisen puun tärkeimmäksi nostama selittävä muuttuja eli alkuosan sanaluokka. Kolmannessa ja viimeisessä puussa on mukana enää yksi selittävä muuttuja eli yhdyssanan merkkimäärä.

Yhdyssanan merkkimäärä on luokitettu kahta ensimmäistä puumallia varten. Ilman luokitusta pienet merkkimäärän muutokset saisivat liian suuren painoarvon näissä puissa (luvun viimeisessä puussa käsittelen merkkimäärää luokittamattomana)⁹⁹. Käytetyt luokat sekä niihin kuuluvien tapausten määrät on esitelty taulukossa 21.

TAULUKKO 21. Yhdyssanojen luokitus merkkimäärän mukaan

Luokka	Merkkimäärä	Esimerkki	Osuus	Määrä
Lyhyt	5–12	<i>hääyö, etäkauppa, annostyynty, petohysteria</i>	57,2 %	12 415
Pitkä	13–26	<i>jogurttibaari, ketsuppilusikka, tulevaisuustaito, turvallisuuskoordinaattori</i>	42,8 %	9 298
Yhteensä			100 %	21 713

Taulukossa 21 kuvatun luokituksen perustana on käytetty aiempaa tutkimusta sekä luvussa 5.1.2 saatuja tietoja suomen kielen uusien yhdyssanojen merkkimääräjakaumasta. Aineiston merkkimäärän mediaani ja keskiarvo ovat kumpikin 13 merkkiä. Luokkien rajan asettamista mediaanin ja keskiarvon alapuolelle 12 ja 13 merkin välille tukee se, että yhdyssanojen produktiivisuus kääntyy laskuun 10–12

⁹⁹ Tässä luvussa menetelmänä käytetyn luokittavan puun etu on, että malliin voidaan sovittaa erilaisia muuttujia – sekä numeerisia että kategorisia (ks. tarkemmin luku 4.4). Merkkimäärä-muuttujaa ei siis tarvitse luokitaa käytetyn menetelmän vuoksi. Luokitus auttaa kuitenkin tekemään mallista havainnollisimman ja keskittymään olennaiseen. Luvussa esiteltyjen luokittavien puiden antamat päätulokset ovat samankaltaisia siitä riippumatta, onko muuttuja luokitettu vai ei.

merkkiä pitkien tapausten muodostaman huipun jälkeen (ks. luku 5.1.2, kuvio 2, s. 105). Lisäksi 12 ja 13 merkin välillä kulkee sen raja, kuinka paljon ihmisen tarkan näön alueelle mahtuu kerralla merkkejä (Hyönä – P. Niemi 1990; ks. myös luku 2.3.3).

Luokittava puumalli, jonka selitettävänä muuttujana on yhdyssanan alkuosan muoto ja selittävinä muuttujina alkuosan sanaluokka, koko muodosteen sanaluokka sekä luokitettu merkkimäärä, osoittaa, että kaikilla selittäville muuttujilla on tilastollisesti merkitseviä yhteyksiä alkuosan muotojakaumaan. Selittävästä muuttujista tärkeimmäksi luokittava puu nostaa alkuosan sanaluokan. Tämä on odotuksenmukaista, sillä alkuosan muodon ja alkuosan sanaluokan välillä on kolme yksi yhteen -riippuvuutta: 1) Kaikki taipumattomiin tai vajaasti taipuviin sanaluokkiin kuuluvat alkuosat ovat myös muodoltaan taipumattomia tai vajaasti taipuvia (esim. *kanssakädellinen, päälleliimattu*). Kuitenkaan kaikki taipumattomat tai vajaasti taipuvat alkuosat eivät kuulu taipumattomiin tai vajaasti taipuviin sanaluokkiin (esim. adjektiivialkuiset *pikkulevy, tuikituore*). 2) Kaikki sanaluokaltaan epäspesifit alkuosat ovat muodoltaan epäitsenäisiä elementtejä (esim. *anarkofeminismi, etäkirurgi*). Samoin kaikki muodoltaan epäitsenäiset elementit ovat sanaluokaltaan epäspesifisiä. 3) Verbialkuisten yhdyssanojen alkuosat ovat yhtä lukuun ottamatta (*istumaannousu*) yhdysosamuotoisia (esim. *asuintorni, kellumahiiri, nuokkusyreeni*). Kaikki yhdysosamuodot eivät kuitenkaan edusta verbiä (esim. *pienkompostori, päivittäisliikenne*).

Kuvio 14 esittelee puumallin, josta on jätetty pois edellä mainitut ainoastaan yhtä alkuosan muotoa käyttävät ryhmät: taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvalla alkuosalla, sanaluokaltaan epäspesifillä alkuosalla sekä verbillä sanalla alkavat yhdyssanat (kuvion $n = 19\ 846$). Alkuosan mahdollisista muodoista epäitsenäiset elementit putoavat samalla mallista pois. Kuten kuviosta nähdään (puukuvaimen lukemiseen on annettu ohjeita luvussa 4.4), kaikki selittävät muuttujat – alkuosan sanaluokka, koko muodosteen sanaluokka ja koko muodosteen pituus – ovat mukana luokittavassa puumallissa. Niillä kaikilla on siis tilastollisesti merkitsevä vaikutus selitettävän muuttujan eli alkuosan muodon jakaumaan. Alkuosan sanaluokka on tässäkin mallissa selittävästä muuttujista tärkein (noodi 1).

Havainnollisuuden vuoksi sekä tässä että muissa luokittavissa puissa selitettävän muuttujan arvot on järjestetty lehtinoodeissa koko aineistossa havaittuun yleisyysjärjestykseen (ks. luku 5).

Luokittavan puun (kuvio 14) mukaan alkuosan muotojakaumaltaan toisistaan eroavat selvimmin a) numeraali-, pronomini- substantiivialkuiset yhdyssanat sekä b) adjektiivialkuiset ja useaa sanaluokkaa edustavalla sanalla alkavat yhdyssanat (noodi 1). Lehtinoodien eli kuvan alalaidassa olevien pienten pylvädiagrammien perusteella alkuosan muotojakaumassa on vähiten vaihtelua numeraali-, pronomini- ja substantiivialkuisien yhdyssanojen ryhmässä (lehtinoodit 4, 5, 8, 9, 11 ja 12). Eniten vaihtelua on adjektiivialkuisien yhdyssanojen ryhmässä (lehtinoodit 16, 18, 19, 21 ja 22). Näiden ryhmien väliin jäävät useaa sanaluokkaa edustavalla sanalla alkavat yhdyssanat (lehtinoodi 23). Oleelliset erot adjektiivialkuisien ja muiden yhdyssanojen välillä ovat seuraavat: 1) yhdysosamuodolla alkavat yhdyssanat ovat yleisempiä adjektiivialkuisien kuin muiden yhdyssanojen ryhmässä, 2) taipumattomilla ja vajaasti taipuvilla sanoilla alkavat yhdyssanat ovat yleisempiä adjektiivialkuisien kuin muiden yhdyssanojen ryhmässä ja 3) nominatiivialkuisuus on adjektiivialkuisien yhdyssanojen ryhmässä harvinaisempaa kuin muiden yhdyssanojen ryhmässä.

Kaikki yllä mainitut seikat ovat pääteltävissä myös sellaisten rakenteellisten seikkojen perusteella, jotka ovat nousseet esiin jo aiemmin tässä tutkimuksessa (ks. varsinkin luvut 5.2.4 ja 5.3.1–5.3.2). Yhdysosamuotojen suurta määrää adjektiivialkuisissa yhdyssanoissa selittää se, että yhdyssanan alkuosana toimivista adjektiiveista huomattava osa on *-nen*-loppuisten sanojen *-s*-loppuisia ja *-i*-loppuisia yhdysosamuotoja (esim. *rinnakkaisvalmiste*, *virtuaalikiikho*). Taipumattomien ja vajaasti taipuvien alkuosien yleisyys adjektiivialkuisien joukossa puolestaan selittyy sillä, että kun taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvat sanat on pudotettu pois luokittavasta puusta (ks. edellä), taipumattomien ja vajaasti taipuvien sanojen ryhmään jää vain joukko tiheäkäyttöisiä adjektiiveja (esim. *pikku-* ja *ensi-*) sekä substantiivit *syys-* ja *kyse-* (*kyseenalaistaja*, *syyskurpitsa*) ja taipumaton indefiniitipronomini *joka-* (esim. *jokanuori*). Nominatiivialkuisia yhdyssanoja on adjektiivialkuisien joukossa suhteellisen vähän siksi, että yhdysosamuotojen ja taipumattomien ja vajaasti taipuvien alkuosien suuri määrä laskee niiden suhteellista osuutta.

Siirryn seuraavaksi tarkastelemaan lähemmin noodista 1 noodin 2 suuntaan lähtevää oksaa, joka kuljettaa mukanaan numeraali-, pronomini-¹⁰⁰ ja substantiivialkuisia yhdyssanoja. Noodin 2 mukaan lyhyiden ja pitkien yhdyssanojen alkuosan muotojakauma eroaa tässä joukossa toisistaan tilastollisesti merkitsevästi. Noodien 3 ja 6 mukaan yhdysadjektiivien alkuosan muotojakauma eroaa muita sanaluokkia edustavien yhdyssanojen alkuosan muotojakaumasta sekä pitkien että lyhyiden tapauksen ryhmässä. Pitkien sanojen ryhmässä myös yhdysubstantiivien

100 Luvuissa 5.2.4 ja 6.2.2 olen huomauttanut, että *moni*-kvanttoripronominilla alkavien yhdysanojen osuus on pronomiinalkuisien yhdyssanojen ryhmässä niin suuri (71,3 %), että ne pääsevät yksinään määräämään ryhmän käyttäytymisestä. Kaikkia pronomineja voidaan kuitenkin käsitellä mallituksessa yhtenä ryhmänä, sillä *moni*-pronominilla ja muulla pronominilla alkavien yhdyssanojen alkuosan muotojakaumassa ei ole tilastollisesti merkitsevää eroa (Fisherin tarkan testin $n = 97$, $p = 0,11$). Testauksessa ovat mukana vain nominatiivilla, genetiivillä ja muulla sijamuodolla alkavat tapaukset, sillä *moni*-pronomini ei voi esiintyä yhdysosamuodossa eikä se ole taipumaton tai vajaasti taipuva kuten osa muista pronomineista. Mikäli nämä tapaukset otetaan mukaan testaukseen, ero ryhmien välillä on tilastollisesti merkitsevä.

sekä yhdysadverbien ja -verbien välillä on tilastollisesti merkitsevä ero (noodi 7). Lisäksi pitkien yhdysadjektiivien ryhmässä numeraalialkuisten yhdyssanojen alkuosan muotojakauma eroaa tilastollisesti merkitsevästi pronomini- ja substantiivialkuisten jakaumasta (noodi 10).

Nominatiivi on selvästi yleisin alkuosan muoto kaikissa tämän oksan alaryhmissä (lehtinoodit 4, 5, 8, 9, 11 ja 12). Kaikkein yleisintä nominatiivialkuisuus on lyhyiden numeraali-, pronomini- ja substantiivialkuisten yhdysadverbien, -substantiivien ja -verbien ryhmässä (noodi 5). Valtaosa tämän ryhmän jäsenistä on substantiivialkuisia substantiiveja (esim. *aiherivi*, *tietomurto*). Yhtä ryhmää lukuun ottamatta seuraavaksi yleisin alkuosan muoto on genetiivi. Genetiivialkuisuus on jonkin verran yleisempää yhdysadjektiivien kuin muita sanaluokkia edustavien yhdyssanojen ryhmissä (vrt. lehtinooideja 4 ja 11 lehtinooideihin 5, 8, 9 ja 11). Pitkien numeraalialkuisten yhdysadjektiivien ryhmässä toiseksi yleisin alkuosan muoto on kuitenkin yhdysosamuoto (noodi 12). Yhdysosamuotojen yleisyys tässä pienessä ryhmässä ($n = 21$) on seurausta *kolme*-numeraalin yhdysosamuodon käytöstä (ryhmässä 4 kpl, esim. *kolmimuunteinen*). Kaiken kaikkiaan yhdysosamuodon käyttö on aineistossa yksinomaista *kolme*- ja *neljä*-numeraaleista (esim. *nelikenttä*).

Seuraavaksi erittelen tarkemmin kuviossa 14 noodista 1 noodiin 13 ja siitä eteenpäin kulkevaa oksaa, joka kuljettaa mukanaan adjektiivialkuisia ja useaa sanaluokkaa edustavalla sanalla alkavia yhdyssanoja. Noodin 13 mukaan näiden kahden ryhmän alkuosan muotojakauma eroaa toisistaan tilastollisesti merkitsevästi.

Adjektiivialkuisten yhdyssanojen alkuosan muotojakauma on erilainen sen mukaan, mitä sanaluokkaa koko muodoste edustaa ja onko kyse pitkästä vai lyhyestä yhdyssanasta. Noodin 14 mukaan selvimmin muista eroavat yhdysadjektiivit. Noodin 15 mukaan adjektiivialkuisten yhdysadverbien, -substantiivien ja -verbien kesken sanaluokkaa tärkeämpi erottava tekijä on sanan pituus. Noodin 17 mukaan pitkien adjektiivialkuisten sanojen ryhmässä toistaan tilastollisesti merkitsevästi eroavat a) yhdyssubstantiivit ja b) yhdysadverbit ja -verbit. Noodin 20 mukaan myös lyhyiden ja pitkien adjektiivialkuisten yhdysadjektiivien alkuosan muotojakaumat eroavat toisistaan tilastollisesti merkitsevästi.

Adjektiivialkuisten yhdyssanojen ryhmissä on vähemmän nominatiivimuotoisia alkuosia kuin muihin sanaluokkiin kuuluvilla sanoilla alkavissa yhdyssanoissa (vertaa noodista 1 noodin 2 suuntaan lähtevää oksaa noodien 13 ja 14 suuntaan lähtevään oksaan). Nominatiivi on kuitenkin yleisin alkuosan muoto myös adjektiivialkuisten yhdyssanojen joukossa yhtä ryhmää lukuun ottamatta. Lehtinoodin 18 mukaan pitkien adjektiivialkuisten yhdyssubstantiivien ryhmässä yleisin on yhdysosamuotoinen alkuosa. Ryhmässä yhdysosamuodoista yleisimmät ovat *-inen*-loppuisten vierasperäisten adjektiivien *-i*-loppuisten yhdysosamuoto (135 kpl, esim. *sosiaalitaksi*, *virtuaalilyhydyke*) ja *-nen*-loppuisten adjektiivien *-s*-loppuinen yhdysosamuoto (125 kpl, esim. *erikoismaito*, *näennäisviljely*). Melko yleisiä ovat myös tiheästi toistuvien *pieni*, *suuri*, *täysi* ja *uusi*-sanojen konsonanttivartalon kaltaiset yhdysosamuodot (yht. 39 kpl, esim. *täyshuntu*, *uusasiällisuus*). Muita yhdysosamuotoja ryhmässä ei esiinny (ks. yhdysosamuotojen jaottelu luvussa 5.3.2.3).

Yhdysosamuotoinen alkuosa on nominatiivin jälkeen yleisin myös muissa adjektiivialkuisten yhdyssanojen ryhmissä yhtä lukuun ottamatta: lehtinoodin 22 mukaan lyhyiden adjektiivialkuisten yhdysadjektiivien pienessä ($n = 38$) ryhmässä nominatiivin jälkeen yleisin on genetiivi (kuten totesin edellä noodista 1 noodin 2 suuntaan kulkevan oksan esittelyn yhteydessä, genetiivialkuisuus on lehtinoodien 4 ja 11 mukaan verrattain yleistä myös muissa pelkkiä yhdysadjektiiveja sisältävissä ryhmissä). Kun vielä huomioidaan, että aineiston genetiivimuotoisella adjektiivilla alkavista muodosteista kaikki ovat adjektiiveja (esim. *ihananturha*, *viiltävänvaalea*), genetiivialkuisuuden voidaan päätellä olevan alkuosan sanaluokan lisäksi vahvasti yhteydessä myös koko muodosteen sanaluokkaan (ks. kuvio 15 jäljempänä sivulla 196; ks. myös Saukkonen 1973: 337–338).

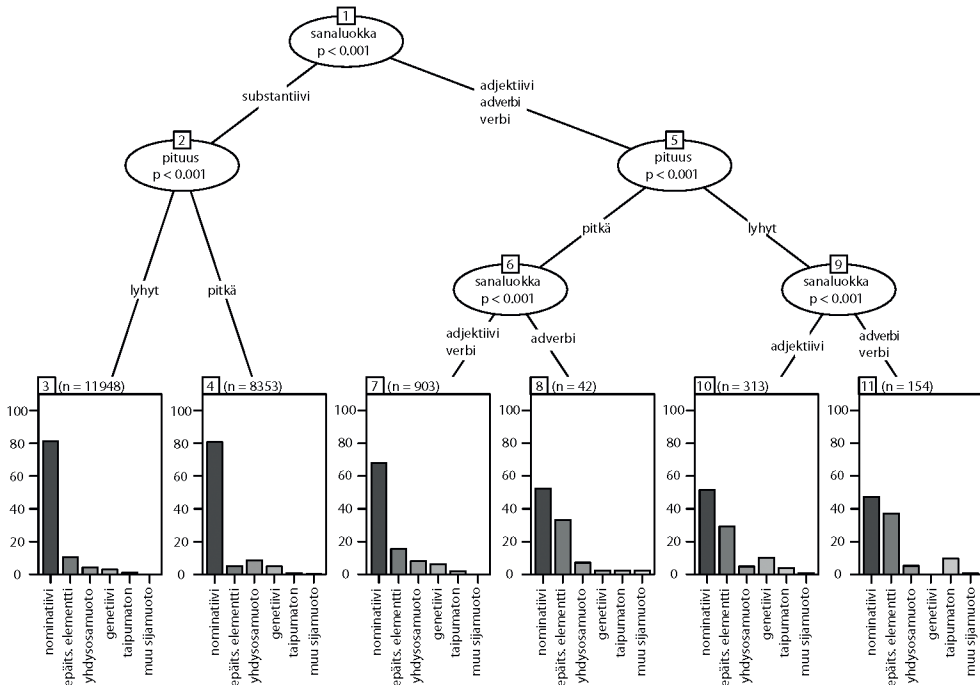
Useaa sanaluokkaa edustavalla sanalla alkavien yhdyssanojen ryhmän sisällä ei ole sanaluokkaan tai muodosteen pituuteen perustuvaa tilastollisesti merkitsevää vaihtelua alkuosan muotojakaumassa (noodi 23). Ryhmässä yleisin alkuosan muoto on nominatiivi ja toiseksi yleisin yhdysosamuoto. Yhdysosamuotojen yleisyys on seurausta kahdesta syystä. Ensinnäkin adjektiivialkuosien tapaan monet useaa sanaluokkaa edustavista alkuosista ovat *-nen*-loppuisia ja esiintyvät siksi yhdyssanoissa *-s*-loppuisessa yhdysosamuodossa (ryhmässä 37 kpl, esim. sekä adjektiivin että substantiivin potentiaalini omaavat alkuosat sanoissa *rikolliskytkentä* ja *venäläismatkailu*). Toiseksi sekä substantiivin että adjektiivin potentiaalini omaavat värinimitykset esiintyvät yhdyssanojen alkuosana usein yhdysosamuotoisina. Varsinkin yleisimmin käytössä olevista värinimityksistä yhdysosamuodot ovat yksinomaisia (ryhmässä yhteensä 93 kpl; esim. *keltapyydys*, *valkosormisuus*).

Kuviota ja erityisesti sen lehtinoojeja tulkittaessa on muistettava ottaa huomioon, että lehtinoodien kuvaamat ryhmät ovat hyvin erikokoisia. Koska lehtinoodit kuvaavat nimenomaan eri alkuosamuotojen osuuksia aineistossa (ei absoluuttisia määriä), pieniä ryhmiä kuvaavissa diagrammeissa hyvin harvinaisetkin muodot voivat nousta esiin kun taas suurimpia ryhmiä kuvaavissa diagrammeissa jopa useat kymmenet tapaukset saattavat kadota suureen massaan. Muulla sijamuodolla alkavia yhdyssanoja on koko aineistossa vain 28. Kaikista ryhmistä eniten muulla sijamuodolla alkavia tapauksia on pitkien substantiivialkuisten yhdysubstantiivien joukossa (11 kpl, esim. *rännäsätallaaja*, *tuppeensahaus*). Puukuviossa nämä sanat ovat pitkien substantiivi- ja pronomiinalkuisten yhdysubstantiivien yhdessä muodostamassa lehtinoodissa 9, mutta eivät näy suuren ryhmän ($n = 7\ 225$) pylväsdiagrammissa lainkaan. Sen sijaan kahdessa pienempiä ryhmiä kuvaavissa lehtinoodissa (noodit 4 ja 8) muulla sijamuodolla alkavat tapaukset nousevat juuri ja juuri näkyviin, vaikka tosiasiaassa niitä kuuluu noodien kuvaamiin ryhmiin vain muutama. Kaiken kaikkiaan muilla sijamuodoilla kuin nominatiivilla ja genetiivillä alkavia yhdyssanoja on aineistossa niin vähän, että niiden jakaumasta on vaikea tehdä tämän enempää luotettavia päätelmiä.

Luokittavasta puusta on helpointa havainnoida sen tärkeimmäksi nostamaa selittävää muuttujaa. Muiden selittävien muuttujien vaikutuksesta selitettävän muuttujan jakaumaan on vaikeampi saada kokonaiskuvaa. Tämän vuoksi ryhdyn

seuraavaksi pilkkomaan puuta tyvestä alkaen. Seuraavassa luokittavassa puussa alkuosan sanaluokan vaikutus on jätetty huomiotta siksi, että koko muodosteen sanaluokan ja pituuden yhteyttä alkuosan muotoon olisi helpompi havainnoida. Kuvion 15 luokittavassa puussa selitettävänä muuttujana on siis alkuosan muoto ja selittävinä muuttujina koko muodosteen sanaluokka ja pituus. Kuviosta 15 nähdään edellistä kuviota 14 havainnollisemmin esimerkiksi jo edellä mainitsemani yhdysadjektiivien ja genetiivialkuosien yhteys.

Tätä ja seuraavaa luokittavaa puuta (kuvio 16) varten olen palauttanut aineistoon kuviosta 14 puuttuneet taipumattomiin ja vajaasti taipuviin sanaluokkiin kuuluvalla alkuosalla ja sanaluokaltaan epäspesifillä alkuosalla sekä verbillä alkavat yhdyssanat (kuvioden $n = 21\ 713$). Kahden seuraavan kuvion lehtinoodeissa näkyy siksi edellisestä kuviosta 14 puuttunut epäitsenäisten elementtien pylväk.



KUVIO 15. Luokittava puu, jossa selitettävänä muuttujana yhdyssanan alkuosan muoto ja selittävinä muuttujina koko muodosteen sanaluokka ja pituus ($n = 21\ 713$, ennustavuus 79,8 %)

Luokittavan puun (kuvio 15) mukaan selvin ero alkuosan muotojakaumassa on a) yhdyssubstantiivien ja b) yhdysadjektiivien, -adverbien ja -verbien välillä (noodi 1). Noodin 5 mukaan yhdysadjektiivien, -adverbien ja -verbien välillä alkuosan muotojakauman kannalta sanaluokkaa tärkeämpi erottava tekijä on koko yhdyssanan pituus. Noodien 6 ja 9 mukaan myös yhdysadjektiivien, -adverbien ja -verbien

alkuosan muotojakaumissa on keskinäisiä eroja, mutta erot ovat erilaisia lyhyiden ja pitkien yhdyssanojen ryhmissä. Lyhyiden sanojen ryhmässä toisista tilastollisesti merkitsevästi eroaa yhdysadjektiivien jakauma (noodi 6), kun taas pitkien sanojen joukossa toisista eroaa yhdysadverbien jakauma (noodi 9).

Kuvion 15 lehtinooodeista nähdään, että eri alkuosan muotojen yleisyysjärjestys on samankaltainen lähes kaikissa ryhmissä koko muodosteen sanaluokasta ja pituudesta riippumatta. Kaikissa ryhmissä yleisin alkuosan muoto on nominatiivi. Sitä seuraavat järjestyksessä epäitsenäiset elementit, yhdysosamuodot ja genetiivi. Vähiten on taipumattomia ja vajaasti taipuvia sekä muussa sijassa kuin nominatiivissa tai genetiivissä olevia alkuosia. Yleisyysjärjestykseen on neljä poikkeusta: 1) Noodin 4 mukaan pitkien yhdyssubstantiivien ryhmässä yhdysosamuotoiset alkuosat ovat yleisempiä kuin epäitsenäiset elementit. 2) Noodin 8 mukaan pitkien yhdysadverbien pienessä ($n = 42$) ryhmässä genetiivillä, taipumattomalla tai vajaasti taipuvalla sanalla ja muulla sijamuodolla alkavien sanojen osuudet ovat yhtä suuret. 3) Noodin 10 mukaan lyhyiden yhdysadjektiivien ryhmässä genetiivialkuosat ovat yleisempiä kuin yhdysosamuodot. 4) Noodin 11 mukaan lyhyiden yhdysadverbien ja -verbien ryhmässä taipumattomat ja vajaasti taipuvat alkuosat ovat yleisempiä kuin yhdysosamuodot ja genetiivit. Niiden ryhmässä ei ole lainkaan genetiivialkuisia tapauksia, ja taipumattomat ja vajaasti taipuvat alkuosat ovat niiden joukossa yleisempiä kuin missään muussa ryhmässä. Kahta *ensi-* ja kahta *eri-*adjektiivin esiintymää lukuun ottamatta kaikki yhdysadverbien ja -verbien taipumattomat ja vajaasti taipuvat alkuosat ovat adverbejä. Tyypillisiä ovat esimerkiksi suuntaa ilmaisevat alkuosat (esim. *poisoppia, ylösmaata*).

Edellä luvussa 5.3.2.4 olen todennut sekä epäitsenäisten alkuosaelementtien että erilaisten yhdysosamuotojen olevan uudessa yhdyssanastossa yleisempiä kuin vanhastaan tyypillisenä pidetyn genetiivialkuisuuden. Luokittava puumalli osoittaa, että kyse ei ole vain yhtä tai kahta yhdyssanaryhmää koskevasta ilmiöstä, vaan genetiivit ovat epäitsenäisiä elementtejä ja yhdysosamuotoja harvinaisempia käytännössä kaikissa sanaluokissa ja yhdyssanan pituudesta riippumatta. Erityisen yleisiä epäitsenäiset elementit ovat kuitenkin yhdysadjektiiveissa, -adverbeissa ja -verbeissa. Genetiivialkuisuus on yleisintä lyhyissä yhdysadjektiiveissa. Genetiivialkuisuuden ja yhdyssanan adjektiiviuuden verrattain vahva yhteys on huomattu myös aiemmassa tutkimuksessa (esim. Saukkonen 1973: 337–338; ks. uusien yhdysadjektiivien ja sanakirja-aineiston vertailu luvusta 5.3.1).

Vähiten nominatiivialkuja ja eniten epäitsenäisiä elementtejä on lyhyissä yhdysadverbeissa ja -verbeissa (noodi 11, esim. *ikiharvoin, pikapaistaa*). Epäitsenäisten elementtien osuus on suuri myös lyhyissä yhdysadjektiiveissa (noodi 10) ja ja pitkien yhdysadverbien pienessä ryhmässä (noodi 8, $n = 42$). Yhdysadjektiiveissa tyypillisiä epäitsenäisiä alkuosaelementtejä ovat elementit, jotka esimerkiksi kieltävät jonkin ominaisuuden tai kertovat ominaisuuden määrästä (esim. *epäspesifi, semihuoleton, supersöpö*). Yhdysverbeissa tyypillisimmät epäitsenäiset alkuosaelementit kertovat, millä tavalla jotakin tehdään (esim. *lähilukea, salalukea*).

Epäitsenäisten elementtien keskittyminen lyhyisiin yhdyssanoihin selittyy niiden omalla lyhyydellä. Palaan yhdyssanan pituuden ja epäitsenäisten alkuosaelementtien esiintymisen yhteyteen vielä seuraavan puukuvaimen yhteydessä (ks. kuvio 16, s. 199).

Luokittavasta puusta ilmi käyvien asioiden lisäksi aiemmassa tutkimuksessa on huomautettu yhdyssanan alussa olevan muun sijamuodon kuin nominatiivin tai genetiivin yhteydestä koko muodosteen sanaluokkaan. ISK:n (§ 403; ks. myös Penttilä 1963 [1957]: § 223, s. 270–271) mukaan muulla sijamuodolla kuin nominatiivilla tai genetiivillä alkavien yhdyssanojen edusosa on yleensä verbikantainen (deverbaali substantiivi tai partisiippi). Tämä pitää paikkansa myös uudessa yhdyssanastossa.¹⁰¹ Mäkisalo (2000a: 120–121) liittää yhdyssanan edusosan deverbaaliuden muiden sijamuotojen lisäksi myös genetiivin runsaampaan esiintymiseen. Hänen mukaansa yhdyssanan deverbaalin edusosan merkitys voi yhdyssanan sisään tunkeutuessaan johtaa määriteosan genetiivimuotoisuuteen (esim. *eränkäynti*, *autonkuljettaja*). Varsinkin monikolliseen tarkoitteeseen viittaavat määriteosat esiintyvät kuitenkin myös nominatiivissa tai konsonanttivartaloisten sanojen kyseessä ollessa yhdysosamuodoissa.

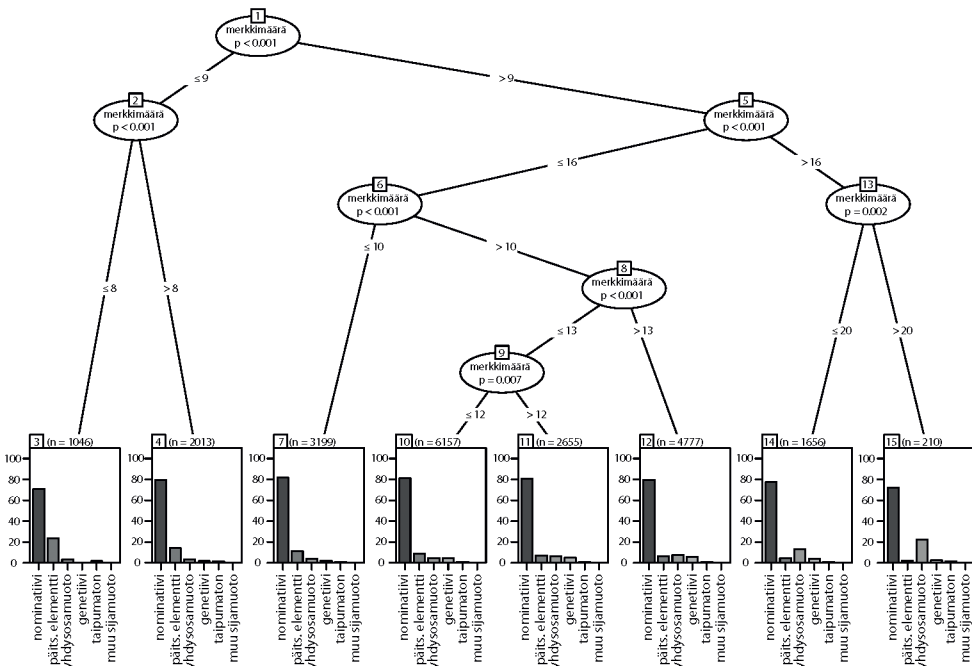
Yhdyssanan osien semanttinen suhde voi olla käytännössä mikä tahansa, eikä määriteosa aina viittaa deverbaalin edussanansa objektiin (esim. *öljylämmittäjä* ei lämmitä öljyä vaan öljyllä). Käytännössä objektinakaan oleva nominatiivisijainen määriteosa ei aina vaadi Mäkisaloon (2000a: 120–121) mainitsemaa monikkotulkintaa (esim. *järjestelmätoimittaja*, *taaperoimittaja*). Ainakin deverbaalit *-jA*-tekijännimijohtimella johdetut jälkiosat esiintyvät kuitenkin myös uudessa yhdyssanastossa tilastollisesti merkitsevästi useammin genetiivialkuosan kanssa kuin substantiiviset perussanat ja denominaalit substantiivit (alkuosan muotojakaumien eroa testaavan Fisherin tarkan testin $n = 1\,000$ ¹⁰², $p < 0,001$ ^{***}, sovitettu standardoitu jäännös $+/-4,4$). Vastaavasti perussanojen ja denominaalien substantiivien yhteydessä alkuosa on useammin nominatiivissa kuin deverbaalien edusosien kanssa (sovitettu standardoitu jäännös $+/-2,2$). Muiden alkuosan muotojen esiintymisessä ei ole ryhmien välillä tilastollisesti merkitseviä eroja.

On hyvä huomata, että genetiivialkuosien yleisyys liittyy nimenomaan yhdyssanan deverbaaleihin jälkiosiin – mutta ei verbijälkiosiin. Myös verbijälkiosien kanssa esiintyvien alkuosien muotojakauma poikkeaa tilastollisesti merkitsevästi substantiivisten perussanojen ja denominaalisten substantiivien kanssa esiintyvien

101 ISK:n (§ 403) mukaan myös adverbi- ja partikkelimääriteosilla on yleensä verbikantainen edusosa. Näin on myös uudessa yhdyssanastossa, vaikka myös adverbin ja denominaalisen sanan yhdistyminen on täysin mahdollista. Tutkimusaineistossa verbikantainen jälkiosa on esimerkiksi sanoissa *jälleenpostitus*, *ristinomistus* ja *sisäänotto*; substantiivikantainen edusosa puolestaan sanoissa *kanssakädellinen*, *poikkipuomi* ja *gineslauntai* 'lauantai, jona ollaan kiinni (armeijassa)'.
102 Testauksessa on käytetty aineistoa, johon valitsin satunnaisesti 500 sellaista yhdyssanaa, joiden jälkiosa on deverbaali *-jA*-tekijännimijohdos (esim. *ammattirähinöitsijä*, *trendinhaistaja*, *käteismaksaja*), ja 500 sellaista yhdyssanaa joiden jälkiosa on joko substantiivinen perussana tai denominaalinen substantiivijohdos (esim. *katuvaate*, *lastenooppera*, *eläkeläispommi*).

alku-osien muotojakaumasta (Fisherin tarkan testin $n = 632^{103}$, $p < 0,001$). Genetiivissä tai muussa sijamuodossa olevien alkuosien esiintymisessä ei kuitenkaan ole tilastollisesti merkitsevää eroa. Sen sijaan verbien yhteydessä yleisempiä ovat epäitsenäiset elementit (sovitettu standardoitu jäännös $+/-6,2$) ja taipumattomat tai vajaasti taipuvat alkuosat (sovitettu standardoitu jäännös $+/-2,9$). Substantiivien yhteydessä puolestaan ovat yleisempiä nominatiivialut (sovitettu standardoitu jäännös $+/-2,9$). Yhdysosamuotojen esiintymisessä ei ole ryhmien välillä merkitsevää eroa.

Seuraavaksi siirryn käsittelemään luokitettavaa puuta (kuvio 16), jossa selittävistä muuttujista on jäljellä ainoastaan yksi: yhdyssanan pituus. Selitettävänä muuttujana on edelleen alkuosan muoto. Tätä puumallia varten aineiston sanojen binäärinen luokitus lyhyisiin (5–12 merkkiä) ja pitkiin (13–26 merkkiä; luokituksesta tarkemmin luvun alussa) yhdyssanoihin on purettu ja selittävänä muuttujana on käytetty numeerista merkkimäärä-muuttujaa. Tämä kuvio havainnollistaa aiempia puukuvaimia (kuviot 14 ja 15) paremmin yhdyssanan pituuden ja alkuosan muotojakauman yhteyttä.



KUVIO 16. Luokitettava puu, jossa selitettävänä muuttujana yhdyssanan alkuosan muoto ja selittävä muuttujana koko muodosteen merkkimäärä ($n = 21\,713$, ennustavuus 79,8 %)

103 Testauksessa mukana kaikki aineiston yhdyssverbis (316 kpl) sekä saman verran satunnaisesti valittuja yhdyssanoja, joiden jälkiosa on joko substantiivinen perussana tai denominaalinen substantiivijohdos.

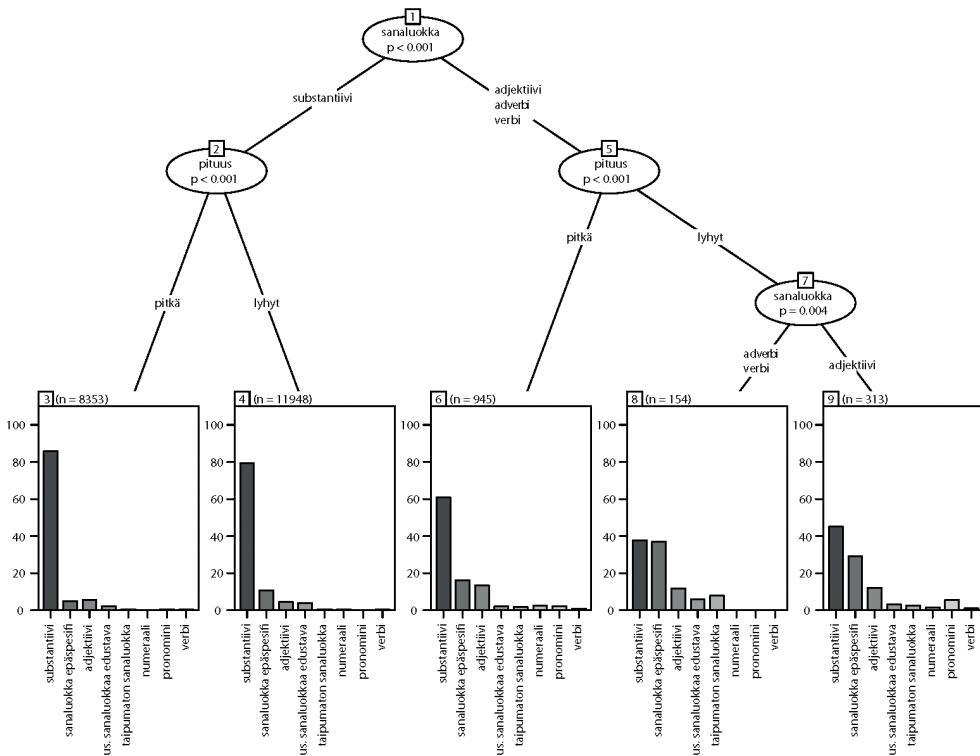
Kuten luokittavasta puusta (kuvio 16) nähdään, nominatiivi on yleisin alkuosan muoto yhdyssanan pituudesta riippumatta. Seuraavina tulevat epäitsenäiset elementit ja yhdysosamuodot, joiden yleisyysjärjestys on kuitenkin erilainen pitkissä ja lyhyissä yhdyssanoissa. 13 tai vähemmän merkkejä sisältävissä sanoissa epäitsenäiset elementit ovat yhdysosamuotoja yleisempiä (lehtinoodit noodista 11 vasemmalle). Epäitsenäisten elementtien suuri osuus lyhimmissä sanoissa on odotuksenmukaista, sillä monet niistä ovat itse varsin lyhyitä. 13 merkin kohdalla alkuosamuotojen yleisyysjärjestys muuttuu, ja sitä pidemmissä yhdyssanoissa yhdysosamuodot ovat epäitsenäisiä alkuosaelementtejä yleisempiä (lehtinoodit noodista 12 oikealle). Lehtinoodin 15 mukaan yli 20 merkkiä sisältävissä yhdyssanoissa myös genetiivialkuisuus nousee epäitsenäisiä elementtejä yleisemmäksi.

Kaikkein lyhimpiä sanoja lukuun ottamatta nominatiivien, epäitsenäisten elementtien ja yhdysosamuotojen jälkeen yleisin alkuosan muoto on genetiivi. 8 merkkiä tai vähemmän sisältävien yhdyssanojen joukossa on vain neljä genetiivillä alkavaa muodostetta (lehtinoodi 3, esim. *ihonsävy*, *puunosto*). Genetiivin harvinaisuus on seurausta siitä, että yleensä kahdesta sanasta muodostuva yhdyssana on jo itsessään pidempi kuin 8 merkkiä, eikä genetiivin päätte näin ollen mahdu näin lyhyisiin yhdyssanoihin. Genetiivialkujen osuus on suurin 10–16 merkkiä sisältävissä yhdyssanoissa (lehtinoodit 10, 11 ja 12). Tätä pidemmissä sanoissa genetiivin osuus jälleen pienenee. Pisimmissä yhdyssanoissa genetiivialkuisuutta suositaan todennäköisesti siksi vähemmän kuin keskipitkissä, että genetiivialkuisuus mahdollistaa usein myös osien erikseen kirjoittamisen ja kielenkäyttäjän tarve ilmauksen osien erilleen kirjoittamiseen on todennäköisesti sitä suurempi, mitä pidemmästä ilmauksesta on kyse.

Kuvion 16 mukaan alkuosan muotojakauma on 12 merkkiin asti sama kuin koko tutkimusaineistossa (ks. myös luku 5.3.1). Pitkissä yhdyssanoissa on eniten poikkeavuuksia yleisestä linjasta myös silloin, kun selitettäväksi muuttujaksi asetetaan koko muodosteen tai alkuosan sanaluokka: Koko muodosteen sanaluokkien yleisyysjärjestys on yhdyssanan pituudesta riippumatta sama kuin tutkimusaineistossa kaiken kaikkiaan (ks. myös luku 5.2.2), mutta yhdysadjektiivien osuus kasvaa ja yhdyssubstantiivien osuus pienenee sen mukaan, mitä pidemmistä yhdyssanoista on kysymys. Alkuosan sanaluokkien yleisyysjärjestys vastaa koko tutkimusaineistosta havaittua järjestystä (ks. myös luku 5.2.4) 15 merkkiin asti. Sen jälkeen adjektiivialkuisten yhdyssanojen osuus nousee sanaluokaltaan epäspesifillä alkuosalla alkavien yhdyssanojen osuutta korkeammaksi. Sekä yhdysadjektiivien että adjektiivialkuosien suurta määrää pitkissä yhdyssanoissa selittää se, että adjektiivit ovat keskimäärin pidempiä kuin useimpiin muihin sanaluokkiin kuuluvat sanat (Niemi-korpi 1990: 88–89; ks. tarkemmin myös luku 7.1.2).

7.2.2 Eri sanaluokkiin kuuluvien sanojen yhdistyminen

Tässä luvussa käsittelemme eri sanaluokkiin kuuluvien sanojen yhdistymistä suomen kielen uusissa yhdyssanoissa. Luokittava puumalli, jonka selitettävänä muuttujana on yhdyssanan alkuosan sanaluokka ja selittävinä muuttujina koko yhdyssanan sanaluokka, alkuosan muoto ja yhdyssanan pituus (luokitettu merkkimäärä, ks. luokituksesta tarkemmin edellinen luku 7.2.1), osoittaa kaikilla selittäville muuttujilla olevan tilastollisesti merkitseviä yhteyksiä alkuosan sanaluokkajakaumaan. Selittävistä muuttujista tärkeimmäksi luokitettava puu nostaa alkuosan muodon, jota seuraa koko muodosteen sanaluokka. Koska olen käsitellyt yhdyssanan alkuosan sanaluokan ja alkuosan muodon yhteyksiä laajasti jo edellisessä luvussa (luku 7.2.1), esittelen tässä luvussa puumallin, jossa alkuosan muodon vaikutus on jätetty huomiotta. Eri sanaluokkiin kuuluvien yhdyssanojen yhdistymismahdollisuudet saadaan näin paremmin esiin. Kuviossa 17 on selitettävänä muuttujana alkuosan sanaluokka ja selittävinä muuttujina koko muodosteen sanaluokka sekä koko muodosteen pituus.



KUVIO 17. Luokittava puu, jossa selitettävänä muuttujana yhdyssanan alkuosan sanaluokka ja selittävinä muuttujina koko muodosteen sanaluokka ja pituus ($n = 21\,713$, ennustavuus 80,2 %)

Kuten luokittavasta puusta (kuvio 17) nähdään, selvimmin muista sanaluokista eroaa yhdyssubstantiivien alkuosan sanaluokkajakauma (noodi 1). Noodin 7 mukaan myös lyhyiden yhdysadverbien ja -verbien alkuosan sanaluokkajakauma eroaa tilastollisesti merkitsevästi lyhyiden yhdysadjektiivien alkuosan sanaluokkajakaumasta. Sen sijaan pitkien sanojen ryhmässä yhdysadjektiivien, -adverbien ja -verbien välillä ei ole tilastollisesti merkitseviä eroja (noodi 6). Luokittavan puun esiin nostamat toisistaan eroavat ryhmät ovat samantyyppiset kuin edellisessä luvussa esitellyssä kuviossa 15 (s. 196), jossa selitettävänä muuttujana oli alkuosan muoto.

Kuvion 17 lehtinoodeista 3 ja 4 nähdään, että varsinkin yhdyssubstantiiveissa ylivoimaisesti yleisimpiä ovat substantiivialkuosat (esim. *maissikoira* 'maissisijauhoista tehtyyn taikinaan leivottu makkara', *lausuntoautomaatti*), joita pitkissä sanoissa seuraavat kaukana perässä adjektiivialkuosat (lehtinoodi 3; *halpaoperaattori*, *varhaisvanhuus*) ja lyhyissä sanoissa sanaluokaltaan epäspesifit alkuosat (lehtinoodi 4; *avopizza*, *minilama*). Sekä pitkissä että lyhyissä yhdyssubstantiiveissa on jonkin verran myös useaan sanaluokkaan kuuluvia alkuosia (*homohäät*, *maalaisyhtytynen*). Vain pieni osa yhdyssubstantiivien alkuosista kuuluu taipumattomiin tai vajaasti taipuviin sanaluokkiin (*kanssakädellinen*, *ulkoahjautuvuus*). Myös numeraali-, pronomini- ja verbialkuosien osuus on pieni (esim. *nollajuusto* 'o % rasvaa sisältävä juusto', *minävaurio*, *kuolinapu*). Yhdyssubstantiivien suurissa ryhmissä pienikin osuus voi kuitenkin merkitä määrällisesti suurempaa joukkoa kuin suuri osuus pienessä ryhmässä.

Yhdysadjektiivien, -adverbien ja -verbien alkuosan sanaluokkajakaumassa on kaiken kaikkiaan enemmän vaihtelua kuin yhdyssubstantiivien alkuosan sanaluokkajakaumassa (lehtinoodit 6, 8 ja 9). Myös näissä sanaluokissa alkuosa on yleisimmin substantiivi, mutta substantiivialkuosien ylivalta ei ole yhtä selvä kuin yhdyssubstantiiveissa (esim. *ajatusohjauksinen*, *videoteitse*, *sohvasurffata*). Muita sanaluokkia edustavissa yhdyssanoissa on paljon varsinkin sanaluokaltaan epäspesifejä alkuosia (esim. *sisäsyöttöinen*, *tasavajaalla*, *täsmäsiivota*). Myös adjektiivialkuosat ovat selvästi yleisempiä muissa yhdyssanoissa kuin yhdyssubstantiiveissa (esim. *tylynhainkeä*, *kovakätisesti*, *ensirekisteröidä*). Lyhyiden yhdysadjektiivien ryhmässä (noodi 9) huomionarvoista on lisäksi pronominalkujen verrattain suuri osuus¹⁰⁴ (esim. *monietninen*, *samakätinen*).

Vähiten substantiivialkuosia on lyhyissä yhdysadverbeissä ja -verbeissä, joissa sanaluokaltaan epäspesifien alkuosien osuus on lähes yhtä suuri kuin substantiivialkuosien (lehtinoodi 8, esim. *pikaviestiä*, *tasavajaalla*). Lyhyiden yhdysadverbien ja -verbien alkuosan sanaluokkajakauma poikkeaa koko aineiston alkuosien sanaluokkajakaumasta myös siinä, että niiden ryhmässä neljänneksi yleisin alkuosan sanaluokka on useaa sanaluokkaa edustavien sanojen sijaan taipumaton tai vajaasti taipuva sanaluokka (*myötäruehoon*, *poisoppia*).

104 *Moni*-pronominilla ja muilla pronomineilla alkavien yhdyssanojen alkuosan sanaluokassa ei ole tilastollisesti merkitsevää eroa (Fisherin tarkan testin $n = 97$, $p = 0,09$). Kuten tässä luvussa käytetyssä osa-aineistossa, testauksessa ovat mukana yhdysadjektiivit, -substantiivit ja -verbit.

7.2.3 Yhteenveto

Yhdyssanan alkuosan muodolla, alkuosan sanaluokalla, koko muodosteen sanaluokalla sekä koko muodosteen pituudella on kaikilla tilastollisesti erittäin merkitseviä yhteyksiä toisiinsa. Alkuosan muodon ja sanaluokan välillä on joitakin ennalta selviä yksi yhteen -riippuvuuksia (esim. kaikki sanaluokaltaan epäspesifit alkuosat ovat muodoltaan epäitsenäisiä elementtejä), mutta muuttujien yhteys pysyy tilastollisesti merkitseväenä vaikka tällaiset tapaukset suljetaan aineiston ulkopuolelle.

Nominatiivialkuisuus on tavallisinta kaikissa yhdyssanoissa niiden alkuosan sanaluokasta riippumatta. Poikkeuksen muodostavat ainoastaan pitkät adjektiivialkuiset yhdyssubstantiivit, joissa alkuosa on yleisimmin yhdysosamuotoinen (tyypillisesti *-nen*-loppuisen sanan *-s*-loppuinen yhdysosamuoto). Kaiken kaikkiaan alkuosan muodossa on eniten vaihtelua adjektiivialkuisissa sanoissa.

Nominatiivi on yleisin alkuosan muoto myös koko yhdyssanan sanaluokasta riippumatta. Eniten nominatiivimuotoisia alkuosia on yhdyssubstantiiveilla. Vähemmän nominatiivialkuja on yhdysadjektiiveilla, *-adverbeilla* ja *-verbeilla*, joissa käytetään huomattavan paljon epäitsenäisiä alkuosaelementtejä. Genetiivimuotoiset alkuosat ovat kaikissa sanaluokissa harvinaisempia kuin epäitsenäiset alkuosaelementit.

Eripituisten yhdyssanojen alkuosien muoto- ja sanaluokkajakaumat sekä koko muodosteen sanaluokkajakaumat ovat samansuuntaisia. Lyhyissä sanoissa käytetään kuitenkin enemmän epäitsenäisiä alkuosaelementtejä kuin pitkissä sanoissa, joissa puolestaan yhdysosamuodot ovat yleisempiä kuin lyhyissä sanoissa. Lisäksi yhdyssanan osana toimivat adjektiivit ovat yleisempiä pitkissä kuin lyhyissä yhdysanoissa. Kaiken kaikkiaan pitkät yhdyssanat poikkeavat lyhyitä enemmän kaikkien yhdyssanojen yhteisistä jakaumista.

Eri sanaluokkia edustavista sanoista yhdyssanoissa yhdistyy tyypillisimmin kaksi substantiivia. Muissa kuin yhdyssubstantiiveissa alkuosien sanaluokkajakaumassa on enemmän vaihtelua. Koko muodosteen sanaluokasta riippumatta alkuosan yleisin sanaluokka on substantiivi, mutta yhdysadjektiivien, *-adverbien* ja *-verbien* alkuosina käytetään myös huomattavan paljon sanaluokaltaan epäspesifejä alkuosia, jotka ovat samalla muodoltaan epäitsenäisiä elementtejä. Myös adjektiivialkuosat ovat selvästi yleisempiä muissa yhdyssanoissa kuin yhdyssubstantiiveissa.

Kaiken kaikkiaan luokittavien puiden avulla tehty analyysi osoittaa, että tyypilliset rakennepiirteet yhdistyvät usein samoihin yhdyssanoihin ja että prototyyppisimmät rakenteet ovat samalla myös produktiivisimpia. Analyysi siis vahvistaa käsitystä, jonka mukaan tyypillinen suomen kielen uusi yhdyssana on nominatiivimuotoisella substantiivilla alkava substantiivi. Uusien yhdyssubstantiivien alkuosan muoto- ja sanaluokkajakauma vastaa hyvin niitä oletuksia, joita aiemman tutkimuksen ja sanaston yleiskuvausten perusteella on mahdollista muodostaa. Samalla analyysi näyttää myös kolikon toisen puolen: yhdyssubstantiiveihin keskittyvät tyypillisten yhdyssanojen kuvaukset eivät ole onnistuneet tuomaan esiin sitä, että

alkuosan muodon ja sanaluokan vaihtelu on suurinta muihin sanaluokkiin kuuluvissa yhdyssanoissa. Poikkeustapaukset esitellään kuvauksissa poikkeustapauksina, mutta vain vähän on kiinnitetty huomiota siihen, että samoihin sanoihin kertyy usein monia epätavallisia rakennepiirteitä (ks. kuitenkin Mäkisalo 2000a: 103–104).

8 Lopuksi

Tässä tutkimuksessa olen selvittänyt, millaiset morfosyntaktiset yhdyssanarakenteet ovat produktiivisia nykyisessä suomen kirjoitetussa yleiskielessä. Produktiivisuuden eri puolista olen käsitellyt nimenomaan rakenteiden suosiota todellisessa kielenkäytössä, en niiden kieliopillista potentiaalia uuden tuottamiseen (luku 3.1.2). Olen lähestynyt aihetta useiden alakysymysten kautta. Olen edennyt tutkimuksessa yksittäisten muuttujien tarkastelusta muuttujien keskinäisiä suhteita tarkasteleviin malleihin. Tutkimusmetodina olen käyttänyt tätä tutkimusta varten luomaani kaksivaiheista metodia, joka koostuu uudissanastossa havaittujen rakenteiden tyyppifrekvenssin tilastollisesta analyysistä sekä varsinkin matalafrekvenssisten rakenteiden kvalitatiivisesta tarkastelusta (ks. luku 3.3.4). Tutkimusaineistona olen käyttänyt toimitetusta painetun median kielestä vuosina 2000–2009 poimittuja uusia ja uudella tavalla käytettyjä yhdyssanoja. Tutkimusaineisto on peräisin Kotimaisten kielten keskuksen Nykysuomen sanastotietokannasta (NSST), jonka pohjalta olen koostanut oman aineistotietokantani (ks. luvut 4.1–4.3).

Tämän luvun ensimmäisessä alaluvussa (8.1) kokoan yhteen tutkimukseni tärkeimmät tulokset ja pohdin niiden syitä. Toisessa alaluvussa (8.2) pohdin tulosten yleistettävyyttä ja arvoa sekä tutkimuksen onnistumista. Luvussa kiinnitän erityistä huomiota tutkimuksessa käytetyn aineiston ja metodin mahdollisuuksiin ja rajoituksiin. Kolmannessa alaluvussa (8.3) esittelen tutkimuksen avaamia jatkotutkimusnäkömiä sekä tutkimuksen tulosten sovellusmahdollisuuksia.

8.1 Mitä tutkimuksen tulokset kertovat?

Tässä luvussa kertaan ensin lyhyesti vastaukset tutkimuksen johdannossa (ks. luku 1.3) esitettyihin kysymyksiin (asetelma 4). Yhteenvedon jälkeen pohdin tarkemmin, mitä saadut vastaukset – yhdessä ja erikseen – tarkoittavat suomen kielen yhdyssanarakenteiden produktiivisuuden kannalta.

ASETELMA 4. Tiivistelmä tutkimuksen johdannossa esitettyihin alakysymyksiin saaduista vastauksista.

Kysymys ja vastaus	Tarkemmin
Kuinka paljon leksikaalisia osia suomen kielen uudet yhdyssanat sisältävät? Valtaosa sisältää kaksi leksikaalista osaa. Myös kolmi- ja neliosaisia yhdyssanoja syntyy jonkin verran, mutta sitä pidemmät uudet yhdyssanat ovat hyvin harvinaisia.	Luku 5.1.1
Kuinka paljon merkkejä suomen kielen uudet yhdyssanat sisältävät? Eniten on 11 merkkiä sisältäviä muodoista. Tätä pidempiä yhdyssanoja on kuitenkin niin paljon, että uusien yhdyssanojen keskipituus nousee noin 13 merkkiin. Merkkimääräjakauksen muistuttaa muuten normaalijakaumaa, mutta siinä on kieliaineistoille tyypillinen pitkä oikealle eli pitkien sanojen suuntaan osoittava häntä. Produktiivisuus kääntyy laskuun 12 merkin jälkeen, mutta jakauma on yhtenäinen aina 33 merkkiin asti.	Luku 5.1.2
Mitä sanaluokkaa suomen kielen uudet yhdyssanat ja niiden jälkiosat edustavat? Valtaosa on substantiiveja. Myös adjektiiveja ja verbejä syntyy jonkin verran, mutta muita sanaluokkia edustavat uudet yhdyssanat ovat harvinaisia. Yhdysadjektiivien osuus on pienempi kuin suomen yleiskielen sanakirjoissa. Koko muodosteen sanaluokka on sama kuin sen jälkiosan, mutta yhdyssanan osien yhdessä luoma mikrokonteksti voi kuitenkin olennaisesti muuttaa jälkiosan sanaluokkapotentiaalia.	Luvut 5.2.2 ja 5.2.3
Mitä sanaluokkaa suomen kielen uusien yhdyssanojen alkuosat edustavat? Valtaosa on substantiiveja. Niiden jälkeen yleisimpiä ovat sanaluokaltaan epäspesifit prefiksinkaltaiset elementit sekä adjektiivialkuosat ja useaa sanaluokkaa edustavat alkuosat. Muihin sanaluokkiin kuuluvat alkuosat ovat harvinaisia.	Luku 5.2.4
Missä muodossa suomen kielen uusien yhdyssanojen alkuosat ovat? Valtaosa on nominatiivissa. Niiden jälkeen yleisimpiä ovat prefiksinkaltaiset epäitse-näiset elementit sekä erilaiset yhdysosamuodot. Genetiivialkuosat ovat vasta neljänneksi yleisimpiä. Muut sijamuodot ovat hyvin harvinaisia. Osa vanhastaan suomen yhdyssanastosta esiintyneistä yhdysosamuodoista ei tuota enää uutta sanastoa produktiivisesti.	Luvut 5.3.1 ja 5.3.2
Taipuvatko suomen kielen uusien yhdyssanojen alkuosat? Alkuosan taipuminen on erittäin harvinaista. Uusia osiltaan kongruoivia yhdyssanoja voi kuitenkin syntyä sanaliitoista tiivistymisen sekä vanhoista sanoista saatavien analogisten mallien kautta.	Luku 5.3.3
Millä tavalla suomen kielen uudet yhdyssanat muodostetaan? Suurin osa suomen kielen uusista yhdyssanoista muodostetaan suoraan kahdesta sanasta yhdistämällä. Kuitenkin yhdysverbit muodostetaan yleisemmin johtamalla aiemmasta yhdyssanasta.	Luku 6.1
Kuinka monipuolisesti suomen kielen uusissa yhdyssanoissa käytetään eri lekseemejä? Eri lekseemejä käytetään hyvin monipuolisesti. Jälkiosissa on enemmän vaihtelua kuin alkuosissa. Vähiten vaihtelua on pronominaalkuisissa ja kotoisella prefiksinkaltaisella alkuosalla alkavissa yhdyssanoissa.	Luku 6.2
Millaisia yhteyksiä suomen kielen uusien yhdyssanojen rakennepiirteiden välillä on? Rakennepiirteiden välillä on useita keskinäisiä vaikutussuhteita. Tärkeimmät havaitut vaikutussuhteet ovat seuraavat: 1) Yhdyssanan leksikaaliset osat ovat sitä lyhyempiä, mitä enemmän niitä on sanassa. Lyhyisiin yhdyssanoihin voidaan valita vapaammin eripituisia osia kuin pitkiin. 2) Merkkimäärällä mitattuna eri sanaluokkia edustavista yhdyssanoista pisimpiä ovat yhdysadjektiivit ja lyhimpiä yhdyssubstantiivit. Osien määrällä mitattuna pisimpiä ovat yhdyssubstantiivit ja lyhimpiä yhdysverbit. Eri sanaluokkiin kuuluvien yhdyssanojen pituuteen vaikuttavat erilaiset syyt. 3) Nominatiivialkuisuus on tavallisinta kaikkiin sanaluokkiin kuuluvissa yhdyssanoissa. Erityisen	Luku 7

<p>yleistä se on kuitenkin yhdyssubstantiiveissa. Yhdysadjektiiveissa, -adverbeissä ja -verbeissä käytetään huomattavan paljon epäitsenäisiä prefiksinkaltaisia alkuosaelementtejä. 4) Substantiivialkuisuus on tavallisinta kaikkiin sanaluokkiin kuuluvissa yhdyssanoissa. Erityisen yleistä se on kuitenkin yhdyssubstantiiveissa. Yhdysadjektiiveissa, -adverbeissä ja -verbeissä käytetään enemmän adjektiivialkuosia kuin yhdyssubstantiiveissa. 5) Epätavalliset rakennepiirteet kertyvät jossain määrin samoihin sanoihin.</p>	
--	--

Asetelmassa 4 usein toistuva sana on *valtaosa*. Kaikista tutkimusaineiston uusista yhdyssanoista noin 55 % on kahdesta substantiivista koostuvia nominatiivialkuisia muodosteita. Kaksiosaisista yhdyssanoista 68,9 % koostuu kahdesta substantiivista, joista ensimmäinen on nominatiivimuotoinen. Varsinkin tätä rakennetta noudattavien yhdyssanojen tyypillinen muodostustapa on suoraan kahdesta sanasta yhdistäminen. Sekä rakenteeltaan että muodostukseltaan tyypillisiä yhdyssanoja ovat esimerkiksi *homekoulu*, *kamerakännykkä*, *kannatuskysely*, *teinipoliitikko* ja *vyöhykeliikenne*.

Kaksi leksikaalista osaa sisältävien yhdyssanojen joukossa seuraavaksi yleisimpiä ovat muodoltaan epäitsenäisellä ja sanaluokaltaan epäspesifillä elementillä (käytännössä pääasiassa prefiksinkaltaisilla elementeillä) alkavat substantiivit (esim. *kaukoleikkaus*, *kyberraivo*). Niitä on noin 7,6 % aineistosta. Seuraavina tulevat genetiivimuotoisella substantiivilla alkavat substantiivit (3,3 %, esim. *brysselinpaini*, *kehonmuokkaus*). Niiden jälkeen keskenään melko tasoissa ovat yhdysosamuotoisella substantiivilla alkavat substantiivit (2,4 %, esim. *asioimistulkki*, *perhosbaari*), nominatiivimuotoisella substantiivilla alkavat adjektiivit (2,2 %, esim. *ammattieettinen*, *hiilineutraali*), nominatiivimuotoisella adjektiivilla alkavat substantiivit (2,2 %, esim. *halpalento*, *omaohjaaja*) ja yhdysosamuotoisella adjektiivilla alkavat substantiivit (2 %, esim. *ilmaisohjelma*, *täyshuntu*). Muita rakenteita esiintyy alle 2 %:ssa tutkimusaineiston 2-osaisista yhdyssanoista.

Nominatiivimuotoisella substantiivilla alkavien yhdyssubstantiivien hyvin selvä määränemmistö aineistossa osoittaa, että suomen kielen uudet yhdyssanat ovat homogeenisempi joukko, kuin kielioppeja ja sanastonkuvauksia lukemalla voisi olettaa. Sanastonkuvauksen ja kielioppien on yleensä tietysti tarkoitus kuvata kaikki se, mikä kielessä on mahdollista, ja marginaaliset rakenteet saavat niissä siksi yleisyyteensä nähden paljon huomiota. Kielenkäytössä toteutuvan rakenteiden produktiivisuuden kuvausta tarvitaankin täydentämään kielioppien ja sanastonkuvauksen antamaa kuvaa kielestä (esim. P. Leino 2002b [1989]: 364–367). Tämä tutkimus on kuitenkin tuonut esiin myös sellaista uutta tietoa, joka on syytä ottaa huomioon tulevia kielenkuvauksia laadittaessa.

Uusista yhdyssanoista *valtaosa* on kaikin tavoin prototyyppisiä yhdyssanoja (prototyyppisin rakenne ei välttämättä aina ole suosituin, ks. esim. Paulsen 2011: 345–365). Aiemmat sanastonkuvaukset eivät perustu kielestä kerättyyn tilastolliseen tietoon, joten kvantitatiivinen vertailu ei ole mahdollista. Näyttää kuitenkin siltä, että suomen kielen uudet yhdyssanat ovat rakenteeltaan aiempaa prototyyppisempiä. Epätyypillisistä piirteistä katoamassa on ainakin osien kongruenssi (ks. luku 5.3.3). Lisäksi muut sijamuodot kuin nominatiivi näyttävät harvinaistuvan

yhdyssanan alkuosina (ks. erityisesti luvut 5.3.1 ja 5.3.2). Kategoriassa on kuitenkin myös yksi varsin selvä uusi murtumakohta: prototyypinen yhdyssana koostuu itsenäisistä lekseemeistä (ks. tarkemmin luku 2.1.2), mutta uudessa yhdyssanas-tossa varsin yleisiä ovat epäitsenäiset prefiksinkaltaiset alkuosat (ks. erityisesti luku 5.3.2.2).

Yksi tutkimuksen keskeisistä huomioista on myös se, että eri rakennepiirteet eivät jakaudu tasaisesti uusien yhdyssanojen joukkoon, vaan epätyypilliset piirteet näyttäivät jossain määrin kertyvän samoihin sanoihin. Esimerkiksi Mäkisalo (2000a: 103–104) on huomauttanut tästä tendenssistä jo aiemmin, mutta suureen aineistoon perustuvaa näyttöä asiasta ei ole ollut. Tämä tutkimus osoittaa, että muihin sanaluokkiin kuuluvien yhdyssanojen alkuosan muodossa ja sanaluokassa on huomattavasti enemmän vaihtelua kuin yhdysubstantiivien alkuosissa (ks. luku 7.2).

Rakenteiden produktiivisuuteen vaikuttavat kaksi vastakkaista voimaa: pyrkimys tuottamisen helpouteen ja pyrkimys ymmärrettävyyteen (vrt. esim. äänne-muutokset). Suomen kielen yleisin yhdyssanarakenne on samalla myös sekä prototyypisin että morfologisesti yksinkertaisin ja siten tuottajan kannalta helpoin mahdollinen. Esimerkiksi lasten on havaittu omaksuvan aikaisemmin ja paremmin sellaiset muodot, joissa morfologinen yhdistäminen perustuu ainoastaan aggluti-naatioon (Nikolaev – Jussi Niemi 2005: 490). Morfologisesti yksinkertaisin rakenne on suomen kielessä myös frekventin, ja lapset omaksuvat sen ensimmäisenä myös siksi (Argus – Johansen Ijäs – Laalo 2014). Vastaanottajan kannalta morfologisesti yksinkertaiset yhdyssanarakenteet ovat kuitenkin vaikeampia. Morfologisesti yksin-kertaisissa yhdyssanarakenteissa tieto on pakattu niin tiiviisti, että uusien ilmausten ymmärtäminen ei yleensä ole mahdollista ilman maailmantietoa ja tietoa kontekstista (useimmat yhdysosamuodot ovat tässä mielessä samankaltaisia nominatiivial-kuosien kanssa, ks. varsinkin luvut 2.3.1 ja 2.3.2). Suomen kielessä yhdyssanoista voidaan tehdä yksiselitteisempiä käyttämällä yhdyssanan alkuosassa muuta sija-muotoa kuin nominatiivia (vrt. esim. *tyvikohottaja* ~ *tyvenkohottaja*; *maamuuttaja* ~ *maahanmuuttaja* ja *maastamuuttaja*; *maamuuttajan* lähde Kulttuurin rikastumi-nen, muut esimerkit TT).

Nominatiivialun on perinteisesti ajateltu vaihtelevan suomen kielen yhdys-sanoissa varsinkin genetiivin kanssa. Uusista yhdyssanoista kuitenkin vain noin 4 % on genetiivialkuisia (ks. luvut 5.3.1–5.3.2). Osuus on pienempi kuin yleiskie-len sanakirjoissa (ks. luku 5.3.2.4). Tiedon perusteella ei kuitenkaan voida sanoa varmasti, syntyykö uusia genetiivialkuisia yhdyssanoja nykyään vähemmän kuin ennen, sillä sanakirjoihin päätyneet yhdyssanat ovat syntyneet hyvin pitkän ajan kuluessa. Lisäksi yhdyssanoista vain osa vakiintuu osaksi sanastoa, eikä kaikkia vakiintuneitakaan yhdyssanoja oteta sanakirjoihin. Sanakirjojen ja muiden kie-lenkuvausten perusteella vaikuttaa kuitenkin siltä, että genetiivialkuisuus on ollut aikaisemmin suositumpaa yhdyssanojen muodostuksessa.

Genetiivialkuisen yhdyssanarakenteen produktiivisuuden todennäköiselle vähenemiselle voidaan esittää ainakin kolme mahdollista syytä. Syistä kaksi ensim-mäistä liittyvät kirjallisen viestinnän lisääntymiseen ja kolmas kielen muutoksen

yleisiin tendensseihin. 1) Koska sanayhtymän alussa oleva genetiivi voidaan usein kirjoittaa erilleen edussanastaan merkityksen juuri muuttumatta, osa kirjoittajista saattaa kirjoittaa sanaliitoiksi sellaisiakin ilmauksia, joita tutkija pitäisi ennemmin yhdyssanoina. Kielenkäyttäjien omiin yhdyssanakäsityksiin nojaavassa suomen kieltä koskevassa tutkimuksessa erikseen kirjoittaminen on kuitenkin katsottava osoitukseksi siitä, että oikeinkirjoitussäännöt tunteva kirjoittaja ei pidä ilmausta merkitykseltään yhdyssanamaisen kiteytyneenä (ks. luku 2.2.1). Toisaalta epävarmuus genetiivialkuisen sanayhtymän oikeinkirjoituksesta voi myös ohjata kirjoittajan valitsemaan mieluummin oikeinkirjoituksen kannalta yksiselitteisemmän nominatiivialkuisen rakenteen. 2) Yhdyssanojen on todettu olevan suomessa tyypillisiä nimenomaan yleiskielelle (Tuomi 1989: 28). Lisäksi yhdyssanat ovat selvästi yleisempiä kirjoitetussa kuin puhutussa kielessä (Pajunen–Palomäki 1984: 64). Yleiskielisen ja kirjallisen viestinnän jatkuvasti lisääntyessä myös yhdyssanojen määrää kasvaa entistä nopeammin. Vaikka genetiivialkuisia yhdyssanoja syntyisi määrällisesti yhtä paljon kuin ennenkin, niiden osuus kaikkien uusien yhdyssanojen joukossa pienee, koska nominatiivialkuisen yhdyssanojen määrä kasvaa ainakin sanakirjaverailujen perusteella aiempaa nopeammin. 3) Rakenteiden yksinkertaistumisen on kielen muutoksen universaali tendenssi. Suomen lisäksi alustavia havaintoja genetiivialkuisen yhdyssanojen produktiivisuuden vähenemisestä on tehty myös virosta, jossa genetiivialkuisuus on vanhastaan ollut nominatiivialkuisuutta yleisempää (keskustelu Reet Kasikin kanssa huhtikuussa 2014; virossa genetiivialkuiset yhdyssanat ovat kuitenkin edelleen enemmistössä, ja myös lapset omaksuvat ne ensimmäisenä, Argus – Johansen Ijäs – Laalo 2014). Sekä suomea että viroa voivat ohjata muutokseen myös käänöslainoja antavista kielistä saadut mallit.

Nominatiivin lisäksi suomen kielen uusien yhdyssanojen alkuosassa genetiiviä yleisempiä ovat myös prefiksinkaltaiset alkuosat. Prefiksinkaltaisten elementtien käyttöä suomen kielen yhdyssanoissa pidetään alkuaan vieraana vaikutuksena, ja todennäköisesti siksi ne ovat jääneet yleisyyteensä nähden vähälle huomiolle kieliopeissa ja sanastonkuvauksissa. Viimeistään tämä tutkimus on kuitenkin osoittanut, että prefiksinkaltaisten elementtien käyttö on otettava suomen kielen sananmuodostusta koskevissa tutkimuksissa ja sanaston yleiskuvauksissa tasavertaisena huomioon muiden sananmuodostuskeinojen kanssa. Tähän mennessä prefiksinkaltaisten elementtien olemusta, käyttöä ja määritelmiä on käsitelty yhdyssanojen muodostusta käsittelevässä tutkimuksessa vain vähän (ks. kuitenkin Häkkinen–Björkqvist 1991: 51–52), ja esimerkiksi ISK:n maininnat prefiksinkaltaisten elementtien käytöstä ovat hajanaisia. Sen lisäksi, että oma tutkimukseni osoittaa prefiksinkaltaisten elementtien käytön keskeisyyden uusien yhdyssanojen muodostuksessa, olen tutkimuksessani myös tarkentanut yhdyssanan alussa esiintyvien prefiksinkaltaisten elementtien määritelmää (ks. luku 5.3.2.2).

Viestintä on yhteistyötä, jossa puhuja ja kuulija luovat merkityksiä yhdessä (ks. esim. Mustajoki 2012). Esimerkiksi Aronoffin (1976: 45) mukaan merkityksen ymmärtämisen helppous voi edistää rakenteen produktiivisuutta. Ilmauksen tuottaminen on kuitenkin aina kognitiivisesti kuormittavampaa kuin viestin

tulkitseminen (esim. Levinson 2009: 29). Lisäksi yksiselitteiset mutta pitkät ilmaukset hidastavat viestintää, joten tieto kannattaa pakata tiiviisti, kunhan se pysyy ymmärrettävänä (Newmeyer 2003: 693). Morfologisesti raskaampien rakenteiden onkin todettu usein häviävän produktiivisuuskamppailuissa keveämmille rakenteille (esim. M. Jantunen 1997: 435). Nominatiivialkuinen rakenne valtaa suomen yhdyssanastossa alaa, koska se on helppo ja yksinkertainen tuottaa mutta samalla riittävän ymmärrettävä: kontekstin apu riittää yleensä ohjaamaan kuulijan oikeaan tulkintaan. Rakenteen produktiivisuutta tukevat myös sen monikäyttöisyys ja aiempi laaja frekvenssi. Tarvittaessa suomen kielessä voidaan kuitenkin edelleen käyttää myös ilmauksen merkitystä tarkemmin rajaavia yhdyssanarakenteita.

Toisin kuin joskus on ajateltu, variaatio tai edes muutokset variaatiossa eivät välttämättä merkitse kielen systeemin muuttumista.¹⁰⁵ Mikäli variaatiossa tapahtuvat muutokset ovat laadullisia – kieleen ilmestyy uusi variantti tai jokin vanha variantti katoaa kielestä kokonaan –, muutokset kuitenkin koskevat todennäköisesti ennemmin tai myöhemmin myös sitä, mitä kielessä pidetään ylipäänsä mahdollisena. Tästä tutkimuksesta ei ole ollut mahdollista rakentaa varsinaista kielen muutoksen kuvausta, koska täysin vertailukelpoista aiempaa tutkimusta ei ole ollut saatavilla. Tutkimuksen tulosten vertailu kielioppeihin, sanastonkuvauksiin ja sanakirja-aineistoihin on kuitenkin nostanut esiin sekä yhden kokonaan aiemmista tutkimuksista puuttuvan yhdyssanarakenteen että useita ilmeisesti produktiivisuuttaan osittain tai jopa kokonaan menettäneitä rakenteita.

Lainasanoista muodostetut *-inen*-loppuiset adjektiivit esiintyvät uusien yhdyssanojen alkuosana toimiessaan *-i*-loppuisessa yhdysosamuodossa (esim. *analogiläheisyys, seksuaaliterveys*). ISK (§ 268) mainitsee lyhyesti, että joidenkin *-inen*-loppuisien vierasperäisten adjektiivien *-nen*-päätteettömät muodot voivat esiintyä adjektiivinkaltaisena yhdyssanan määräiteosana, mutta rakennetta ei ole huomioitu sen enempää ISK:ssa kuin muussakaan aiemmassa tutkimuksessa varsinaisena yhdysosamuotona. Periaatteessa *-inen*-loppuiset vierasperäiset adjektiivit voivat esiintyä yhdyssanan alkuosana *-s*-loppuisessa muodossa siinä missä kotoiset *-nen*-loppuiset sanatkin. Tämän tutkimuksen aineistossa *-i*-loppuisen muodon käyttö on kuitenkin yksinomaista, eikä *-s*-loppuisia yhdysosamuotoja esiinny vierassanoista lainkaan (ks. tarkemmin luku 5.3.2.3). Vierassanoista käytettävien yhdysosamuotojen esiintymisympäristöt kaipaavat vielä tarkempaa tutkimusta, mutta tämän tutkimuksen perusteella on selvää, että *-i*-loppuinen muoto on syytä ottaa jatkossa paremmin huomioon.

Kaikki muut tässä tutkimuksessa produktiivisiksi todetut suomen yhdyssanarakenteet on mainittu myös aiemmissa yhdyssanoja käsittelevissä tutkimuksissa ja sanaston yleiskuvauksissa. Lisäksi niissä on esitelty myös sellaisia rakenteita, jotka eivät tämän tutkimuksen valossa näytä olevan ainakaan enää uudessa sanastossa produktiivisia.

¹⁰⁵ Variaatio ja siinä tapahtuvat muutokset toki ovat kielen muuttumisen edellytyksiä. Kielen muutoksesta ks. Kurki (2005: 24–30).

Selvimmän produktiivisuutensa menettäneisiin kuuluu aiemmassakin tutkimuksessa vanhahtavaksi ja murteelliseksi mainittu monikkovartalon kaltaisten *-i*- ja *-o*-loppuisten yhdyssanamuotojen tyyppi (esim. *lehmihaka*, *aitovieri*). Periaatteessa rakenteen produktiivisuutta rajaa yksittäisten sanojen esiintyminen uudessa yhdyssanastossa, sillä rakenne esiintyy vain tiettyjen sanojen yhteydessä (esim. *jalka*, *lehmä*, *sika*; ks. myös ISK § 416). Koska *-i*- ja *-o*-loppuiset muodot vaihtelevat nominatiivin kanssa, näidenkään sanojen käyttö ei suoraan takaa rakenteen säilymistä produktiivisena. Tämän tutkimuksen aineistossa rakenne esiintyy vain kaksi kertaa (*sikopaja* ja *jalkorätti*; ks. tarkemmin luku 5.3.2.3). Molemmat esiintymät voidaan katsoa vanhasta kielestä uudelleen käyttöön otetuiksi ja sitä kautta aineistoon päätyneiksi. Myös molempien esiintymien tyyli on vanhahtava. Nominatiivimuotoiset alkuosat *jalka*- ja *sika*- esiintyvät aineistossa yhteensä 22 kertaa. Lisäksi muut periaatteessa *-i*- tai *-o*-loppuisessa muodossakin mahdolliset sanat esiintyvät aineistossa vain nominatiivimuodossa (tässä tutkimuksessa omaksi ryhmäkseen erotettuja lukusanojen yhdyssosamuotoja lukuun ottamatta, ks. myös ISK § 416 ja luku 5.3.2.3). Useimmissa uusissa yhdyssanoissa *-i*- tai *-o*-loppuinen muoto tuntuisi ainakin jossain määrin kielenvastaiselta (esim. *sikainfluenssa* [TT] ~ ?*sikoinfluenssa*) tai vähintään vanhahtavalta (esim. *lehmälotto* [TT] ~ ?*lehmilotto*). Alkuosan eri muotojen välillä voi olla myös merkitysero (vrt. *sikatauti* 'sioilla esiintyvä tauti' [TT] ja *sikotauti*).

Myös osaa muista yhdyssosamuodoista käytetään vain tietyistä sanoista (ks. tarkemmin luku 5.3.2.3; ks. myös ISK § 416–417). Esimerkiksi *s*-linkityselementin (esim. *herraskeppi*) sisältävän muodon produktiivisuuden katoamisesta kielii lisäksi vanhahtava tyyliarvo sekä rakenteen harvinaisuus uudessa sanastossa verrattuna samoista alkuosaleksemeistä esiintyviin nominatiivimuotoihin. Tosin *pappi*-alkuosan kanssa *s*-elementin sisältävä muoto on tutkimusaineistossa yksinomainen (3 tapausta, esim. *pappisdiktatuuri*).

Produktiivisuuden kannalta hieman toisentyyppinen alkuosan muoto on muu (kuin *-nen*-loppuisten sanojen *-s*-loppuinen) konsonanttivartalo (esim. *pienkompostori*, *uusnoita*). Myös sitä käytetään yhdyssanan alkuosana vain tietyistä sanoista (ks. tarkemmin luku 5.3.2.3; ks. myös ISK § 416). Osa näistäkin muodoista vaihtelee nominatiivin kanssa, mutta konsonanttivartalon käyttö on uudessa sanastossa kuitenkin selvästi nominatiivia yleisempää. Lisäksi nominatiivia esiintyy vain sanaliitoista tiivistyneissä *-inen*-loppuisissa adjektiiveissa (*pienikustannuksinen*, *suuriso-luinen*). Muulla konsonanttivartalolla alkavia uusia yhdyssanoja syntyy jatkuvasti. Rakennetta ei kuitenkaan voida pitää suoranaisesti produktiivisena rakenteellisella tasolla, sillä se esiintyy vain pienen alkuosalekseemien joukon yhteydessä. Suurin osa näistä leksemeistä on *-i*-loppuisia kotoperäisiä lyhyitä adjektiiveja (poikkeuksia esim. *perse* ja *toinen*). Niiden konsonanttivartaloa vastaava on voinut vakiintua yhdyssanakäyttöön myös siksi, että varsinkin *i*:n loppuheitto ylipäänsä on vanhahtaan ollut suomen murteissa tavallista (Nikkilä 1994: 6–8).

Kieliopissa ja kielenkuvauksissa yleensä esitellyistä rakenteista nykyään epäproduktiivinen on myös rakenne, jossa yhdyssanan osat kongruoivat keskenään

(esim. *omatunto* : *omantunnon* ~ *omatunnon*; ks. tarkemmin luku 5.3.3). Rakennetta voidaan pitää epäproduktiivisena kahdesta syystä. Ensinnäkin rakennetta noudattavat uudet sanat ovat hyvin harvinaisia (tämän tutkimuksen aineistossa vain 6 valinnaisesti kongruoivaa tapausta). Toiseksi kaikille rakennetta noudattaville uusille sanoille on osoitettavissa jokin muu syntytapa kuin suora kahdesta sanasta yhdistäminen. Myös ISK:ssa (§ 421) mainitaan, että uudet sanat asettuvat tähän rakennetyyppiin harvoin, joskin uusia rakennetta noudattavia sanoja voi syntyä sanaliitoista tiivistymisen kautta. Kongruenssin on havaittu olevan väistymässä myös monista vanhastaan kongruoineista yhdyssanoista (R. Lehtinen 1984; Eronen 1996), ja osiltaan kongruoiva yhdyssanarakenne saattaakin olla katoamassa suomen kielestä vähitellen kokonaan. Vaikka tämän tutkimuksen aineisto ei sisällä nimistöä, huomio on mielenkiintoinen myös paikannimiin liittyvän, kielenhuollossa tuon tuosta kysymyksiä herättävän paikannimien ja sukunimien alkuosan taituttamisen kannalta (esim. *Uusimaa* : *Uusimaata* ~ *Uuttamaata*; tarkemmin esim. Sinisalo 1987; Mikkonen 2002).

Osiltaan kongruoivien yhdyssanojen tapaan myös suurin osa muulla sijalla kuin nominatiivilla tai genetiivillä alkavista yhdyssanoista syntyy sanaliitoista tiivistymällä (esim. *käyttöönotto*, *maallemuuttaja*; ks. tarkemmin luku 6.1.2). Muulla sijalla alkavat yhdyssanat ovat uudessa sanastossa hyvin harvinaisia. Rakenne kuitenkin täyttää sellaisia ilmaisutarpeita, joiden vuoksi sen ei voi ennustaa katoavan suomen kielen uudissanoista kokonaan ainakaan vielä. Poikkeuksellisten muodostustapojen vuoksi muulla sijalla alkavia yhdyssanoja ei kuitenkaan voida nähdäkseeni (enää) pitää rakenteellisella tasolla produktiivisena yhdyssanatyypinä suomen kielessä.

Suurin osa sekä Suomessa että muualla tehdystä yhdyssanatutkimuksesta keskittyy yhdyssubstantiiveihin, joita valtaosa sekä uusista että vanhoista yhdyssanoista on. Substantiivit ovat ylipäänsä kielen yleisimpiä sanoja, mutta yhdyssubstantiivien nopealle lisääntymiselle on myös muita syitä. Aiemmissä tutkimuksissa (esim. Haarala 1981: 18; Elomaa 1984: 8; Pentikäinen 1999: 262; Tyysteri 2009: 86, 114) monet erikoiskielet on todettu yleiskieltä substantiivivoittoisemmiksi. Varsinkin alakategorioita kuvaavat luokittavat yhdyssubstantiivit ovat tyyppillisiä erikoiskielille. Maailman teknistyessä ja erikoistuneen tiedon popularisoinnin tarpeen lisääntyessä myös yleiskielessä tarvitaan yhä enemmän alakäsitteiden nimityksiksi sopivia yhdyssubstantiiveja (ks. myös Niemikorpi 1996: 12; Tyysteri 2009: 9–16). Yhdyssubstantiivien runsautta selittää myös se, että samaan sanaperheeseen kuuluu yleensä enemmän substantiiveja kuin muihin sanaluokkiin kuuluvia sanoja. Samasta yhdyssanasta johdettujen johdosyhdyssanojen joukossa voi olla useita erilaisia substantiiveja mutta usein vain yksi tai kaksi verbiä ja/tai adjektiivia (esim. TT:ssä *hohtokeila*, *hohtokeilaaja*, *hohtokeilailu* ja *hohtokeilaus* mutta vain *hohtokeilata*). Erilaisille samasta kannasta muodostetuille yhdyssubstantiiveille on siis enemmän tarvetta kuin yhdysadjektiiveille ja -verbeille.

Yhdyssubstantiiveista saatuja tuloksia ei tule yleistää kaikkiin yhdyssanoihin, sillä ainakin suomen kielessä erilaiset poikkeavat rakennepiirteet keskittyvät erityisesti

muuta sanaluokkia edustaviin yhdyssanoihin. Yhdyssanat ovat tärkeä ilmaisukeino myös muissa sanaluokissa, ja myös muihin sanaluokkiin kuuluvia yhdyssanoja olisi syytä tutkia. Tässä tutkimuksessa esiin noussut mielenkiintoinen mutta vaille selitystä jäänyt seikka on esimerkiksi se, että yhdysadjektiivien osuus suomen kielen uusissa yhdyssanoissa on selvästi pienempi kuin sanakirja-aineistossa.

Eri sanaluokkiin kuuluvat yhdyssanat eroavat toisistaan myös muodostukseltaan (ks. tarkemmin luku 6.1). Suurin osa suomeen nykyään syntyvistä uusista yhdyssanoista muodostetaan yhdistämällä kaksi itsenäistä sanaa, mutta yhdysverbeistä suurin osa on muodostettu aiemmista yhdyssanoista johtamalla. Yhdysverbejä syntyy kuitenkin melko paljon myös yhdistämällä. Kaiken kaikkiaan yhdysverbit näyttävät olevan olennainen osa suomen kielen yhdyssanajärjestelmää. Häkkisen (1987: 19) ja Vaittisen (2003: 47) esittelemistä yhdysverbien historiallisista kerrostuksista edelleen produktiivisina voidaan pitää ainakin adverbi- ja nominatiivialkuisia yhdysverbejä (esim. *uudelleenkouluuttaa*, *piilonaulata*). Sen sijaan adjektiivikantaiset johtamalla muodostetut yhdysverbit (esim. *kansanomainen* > *kansanomaistaa*) näyttävät menettäneen produktiivisuuttaan, ja yhdysverbin kantana onkin uusimassa kielessä tavallisesti yhdyssubstantiivi (esim. *moottorisaha* > *moottorisahata*; vrt. ISK § 430). Uutena ryhmänä produktiiviseksi on katsottava prefiksinkaltaisella elementillä alkavat yhdysverbit (esim. *etäopettaa*, *täsmäkitkeä*).

Yhdysverbien lisääntymisestä suomen kielessä on puhuttu jo pitkään. Tämänkin tutkimuksen perusteella näyttää siltä, että yhdysverbejä syntyy kieleen jatkuvasti lisää ja että niitä käytetään ahkerasti myös tilapäismuodosteina. Koska aiempi tilastollinen vertailukohta puuttuu, tämän tutkimuksen perusteella ei kuitenkaan voida sanoa, lisääntyvätkö yhdysverbit nyt nopeammin kuin ennen. Yhdysverbien lisääntymisen mahdolliseksi syiksi on aiemmassa tutkimuksessa ehdotettu esimerkiksi käänöslainauksesta (esim. Saarimaa 1971 [1947]: 90; O. Nuutinen 1989: 119; tosin ks. myös Rahtu 1984: 419) sekä useita erilaisia muodostusmahdollisuuksia (Levonen 2006: 34, 39–40). Yhdysverbit täyttävät myös monenlaisia ilmaisutarpeita: Toisaalta monet niistä sopivat tiiviytensä vuoksi hyvin varsinkin termeiksi (Vesikansa 1989: 257–258; ks. myös Häkkinen 1987: 21–23), joita käytetään nykyään paljon myös yleiskielessä. Toisaalta varsinkin monet tilapäiset yhdysverbit ovat tyyliltään leikkilisiä (ISK § 431; Levonen 2006: 36–37). Tyylivärinsä vuoksi yhdysverbit sopivat hyvin esimerkiksi pakinoihin ja kolumneihin. Kolumnistien nokkela kielenkäyttö on omiaan innostamaan yhdysverbien käyttöön myös muissa yhteyksissä.

Eri sanaluokkiin kuuluvien sanojen yhdistymisessä on suomen kielessä kielioipillisia rajoituksia ainoastaan verbien kohdalla (yhdyssanan alkuosa ei voi olla persoonamuotoinen verbi), ja yhdistymismahdollisuuksia kuvataankin yleensä suomessa hyvin vapaiksi. Tämän tutkimuksen perusteella nämä mahdollisuudet vaikuttavat vielä laajemmilta, kuin aiemmin on esitetty (vrt. ISK § 407; Jussi Niemi 2009: 244–245; ks. myös Häkkinen 1990: 147–148). Tämän tutkimuksen aineistossa esiintyviä mutta aiemmassa tutkimuksessa mainitsematta jääneitä ovat esimerkiksi pronominin ja substantiivin (esim. *minävaurio*, *samakätisyys*) sekä numeraalin ja verbin (esim. *kaksinapaistaa*, *nelikaistaistaa*) yhdistymät. Kovin yleisiä nämä

rakenteet eivät kuitenkaan ole, ja suurin osa niitä noudattavista yhdyssanoista on aiemmista yhdyssanoista johtamalla muodostettuja. Niitä ei siis välttämättä voida pitää produktiivisina yhdyssanarakenteina, vaikka niitä noudattavia uusia sanoja kieleen syntyykin. Kaiken kaikkiaan ainakin pronominalkuosien produktiivisyyden on suhtauduttava varauksella, sillä valtaosa niistä on *moni*-kvanttoripronominin esiintymiä (ks. tarkemmin esim. luku 5.2.4).

Suomen kielen uusissa yhdyssanoissa yhdistyy yleisimmin kaksi substantiivia. Substantiivit voivat yhdyssanan alkuosassa edustaa myös verbien ja adjektiivien merkityksiä, ja varsinkin verbaalisissa merkityksissä käytetäänkin yhdyssanoissa usein verbin substantiivijohdoksia (esim. *teon-* ja *tekijännimijohdot* senkaltaisissa sanoissa kuin *kokoontumiskatos*, *uuvutuslakko*, *liuottajamikrobi*). Tutkimuksessa havaittu tendenssi epätyypillisten piirteiden kertymisestä samoihin sanoihin näkyy myös eri sanaluokkien yhdistymisessä: muita sanaluokkia edustavat sanat yhdistyvät substantiiveja useammin muiden kuin substantiivien kanssa (luku 7.2.2).

Suurin osa suomen kielen uusista yhdyssanoista on melko lyhyitä. Uusien yhdyssanojen keskipituus on 13 merkkiä. Eniten on kuitenkin 10–12 merkkiä sisältäviä sanoja, joita on noin 36 % kaikista aineiston sanoista (ks. tarkemmin luku 5.1.2). Koska yhdyssanojen muodostaminen on rekursiivista, yhdyssanoissa voi olla periaatteessa loputon määrä leksikaalisia osia. Aiemmin on ajateltu, että käytännössä osia on kuitenkin harvoin enemmän kuin viisi (ISK § 405). Tämä tutkimus osoittaa, että myös viisiosaiset yhdyssanat ovat uudessa sanastossa hyvin harvinaisia (ks. tarkemmin luku 5.1.1). Tärkein syy näin pitkien yhdyssanojen harvinaisuuteen on niiden epäkäytännöllisyys. Eri sanaluokkia edustavista sanoista eniten pitkiä muodosteita on yhdyssubstantiiveissa. Niissä pituus ei ilmeisesti ole yhtä häiritsevää kuin muita sanaluokkia edustavissa yhdyssanoissa, jotka ovat epätyypillisiä jo sanaluokkansa puolesta.

Aiemmissä tutkimuksissa on esitetty pitkien yhdyssanojen lisääntyneen varsinkin erikoiskielissä 1900-luvulta alkaen (Rajakallio 1995: 38; Niemikorpi 1996: 9; varsinkin virkakieltä on syytetty pitkistä ja hankalista yhdyssanoista, Iisa–Piehl 1992: 79–81). Erikoiskielen yhdyssanojen on myös todettu olevan keskimäärin yleiskielen yhdyssanoja pidempiä (Tyysteri 2009: 114). Maailman teknistyessä ja alakategorioiden nimityksiksi sopivien sanojen tarpeen lisääntyessä pitkien yhdyssanojen määrän voisi olettaa kasvavan myös yleiskielessä. Toisaalta epävarmuus oikeinkirjoituksesta voi saada ihmiset välttämään pitkiä yhdyssanoja. Pitkiä yhdyssanoja pidetään yleensä myös tyylillisesti huonoina. Lisäksi esimerkiksi termityössä pyritään yleensä aivan tietoisesti välttämään kovin pitkiä yhdyssanoja (TSK: Millä perusteella termisuosituksia; ks. myös Birdlife: Periaatteet). Koska tälle tutkimukselle sopiva vertailukohta puuttuu toistaiseksi, tilastollista näyttöä muutoksista suomen kielen yhdyssanojen pituudessa ei toistaiseksi ole. Tämän tutkimuksen perusteella voidaan kuitenkin todeta, että suomen kirjoitetun yleiskielen uudet yhdyssanat eivät itse asiassa ole pitkiä.

8.2 Tulosten yleistettävyys ja arvo

Kieliopin ja kielenkäytön vuorovaikutus on kaksisuuntaista. Toisaalta kielioppi ja kieliyhteisön yhteiset normit ohjaavat kielenkäyttöä. Toisaalta kielioppi ja kaikki siihen kuuluvat rakenteet ovat käytön motivoimia – niitä tarvitaan johonkin. Kieliopissa saman asian ilmaisemiselle ja saman rakenteen tulkinnalle on yleensä useita eri vaihtoehtoja. Siihen, minkä rakenteen tai tulkinnan ihmiset todennäköisesti valitsevat, vaikuttavat muiden seikkojen ohella heidän aiemmassa kielenkäytössä havaitsemansa frekvenssit (rakenteen käyttöä koskevat todennäköisyydet on toisinaan nähty myös osana kielioppia, esim. Manning 2002; Clark 2005). Esimerkiksi William Croftin (2000: 120) mukaan laajoista korpuksista kerätty tilastotieto paljastaa kielestä sellaisia asioita, jotka pitäisi huomioida kaikissa variaatiota, käyttöä ja kielen muutosta koskevissa tutkimuksissa. Koska vain kielen käyttöä tarkastelemalla saadaan kuva todellisesta kielestä, tilastotietoa on tärkeää hyödyntää myös kielioppia koskevissa tutkimuksissa. (Ks. myös Clark 2005; Laury-Ono 2005; Bybee 2006.)

Tässä luvussa pohdin tutkimukseni tulosten yleistettävyttä ja arvoa sekä arvioin tutkimuksessa käytettyä tutkimusmetodia. Kuten analyysiosia (luvut 5–7) ja edellinen luku (8.1) osoittavat, tämä tutkimus on tuonut esiin paljon sellaista tietoa, joka on syytä ottaa huomioon tulevissa suomen kielen kieliopin ja sananmuodotuksen kuvauksissa. Tutkimus on vahvistanut oikeiksi monia sellaisia aiemmassa tutkimuksessa esitettyjä näkemyksiä, jotka eivät ole perustuneet suoraan kielenkäyttöä koskevaan tutkimukselliseen näyttöön. Lisäksi tutkimus on osoittanut vääräksi joitain aiemmissa kirjoituksissa esitettyjä arvioita ja nostanut esiin sellaisia seikkoja, joita ei ole aiemmassa tutkimuksessa huomioitu lainkaan. Uuden tiedon tuottamisen lisäksi olen tässä tutkimuksessa soveltanut kansainvälisessä tutkimuksessa tuotettua yhdyssanojen rakennetta koskevaa teoriaa suomen kielen yhdyssanoihin. Tämä tutkimus on myös tähän mennessä laajin produktiivisuutta käsittelevä fennistinen tutkimus. Fennistien lisäksi tämä tutkimus tarjoaa tärkeää tietoa myös typologeille. Typologian kannalta suomen kielen tutkimus on tärkeää muun muassa siksi, että yhdyssanoja koskevien typologisten mallien tueksi saadaan lisää tietoa muista kuin indoeurooppalaisista kielistä (ks. myös Bauer 2009b: 346–347).

Tämä tutkimus antaa hyvän yleiskuvan suomen kirjoitetun yleiskielen uusista yhdyssanoista. Koska tutkimus on ensimmäinen laatuaan, kaikille yhteinen kielimuoto on luonteva valinta tutkimuksen kohteeksi. Lisäksi myös suomen kaltaisista verrattain suurista kielistä kirjoitetut kieliopit kuvaavat yleensä pääasiassa kirjoitettua yleiskieltä. Tutkimuksessa käytetty aineisto sisältää toimitetusta painetun median kielestä vuosina 2000–2009 poimittua uutta sanastoa, eikä tutkimuksen tuloksia voida suoraan yleistää koskemaan muita aineistoja. Myös kirjoitetun yleiskielen alalajien osalta tutkimuksen tuloksia on pidettävä ainoastaan suuntaantavina, koska saman sanomalehden eri osastojen kielen on todettu eroavan toisistaan tilastollisesti (ks. esim. Niemikorpi 1990: 25–39, 1991). Tutkimuksen tulosten perusteella voidaan kuitenkin esittää valistuneita arvauksia muista kielimuodoista

ja aikakaudesta. Tutkimuksen tuloksia voidaan tulevaisuudessa käyttää myös hypoteeseina ja vertailukohtina muita kielimuotoja ja aikakausia koskevissa tutkimuksissa.

Tämä tutkimus on luonteeltaan synkroninen ja kuvaa aikakautta vuodesta 2000 vuoteen 2009. Tutkimuksen ei ole tarkoitus kuvata kielen muuttumista diakronisesta näkökulmasta. Kuten menneisyyden ymmärtäminen aina, vertailu muihin aikakausiin ja aineistoihin kuitenkin tukee nykykieltä koskevan käsityksen luomista. Koska muunlaista aiempia kielimuotoja kuvaavaa tutkimustietoa ei ole tätä tutkimusta tehdessä ollut saatavilla, olen käyttänyt vertailukohtana lähinnä sanakirja-aineistoa tarkastelevia tutkimuksia. Vertailut kertovat ensisijaisesti sanakirjojen ja todellisessa kielenkäytössä tavattavien uusien yhdyssanojen eroista. Koska sanakirjoihin otetut yhdyssanat ovat vanhempia kuin tämän tutkimuksen aineistossa olevat, vertailujen voidaan ajatella heijastavan myös eri aikakausien välillä olevia eroja. Vertailuja tulkittaessa on kuitenkin muistettava ottaa huomioon yhdyssanojen erityisasema sanakirjoissa (läheskään kaikkia vakiintuneitakaan yhdyssanoja ei voida ottaa sanakirjoihin) sekä se, että sanakirjoissa olevat sanat ovat syntyneet pitkän ajan kuluessa. Lisäksi sanakirjoilla on myös normatiivisia tavoitteita, joten ne heijastavat myös sitä, mitä kielessä pidetään toivottavana. Sanakirja-aineiston ja tämän tutkimuksen aineiston välillä tehtyjen vertailujen perusteella ei näin ollen voida sanoa, että kielessä olisi ehdottomasti tapahtunut muutoksia. Myöskään vertailut aiempiin kielioppeihin ja sanastonkuvauksiin eivät ole varma osoitus kielessä tapahtuneista muutoksista, sillä kielioppien ja sanastonkuvauksen tarkoitus on yleensä kuvata kaikki se, mikä kielessä on mahdollista tai toivottavaa. Sekä vertailut sanakirja-aineistoon että kielenkuvauksiin osoittavat kuitenkin mahdollisia kielen muutossuuntia, joihin on jatkossa syytä kiinnittää huomiota.

Meneillään olevien prosessien tavoittaminen on tutkimuksessa aina vaikeaa. Koska tämäkin tutkimus kuvaa jo ohitettua aikaa, tutkimuksen tuloksia ei voida suoraan yleistää koskemaan suomen kieleen juuri tällä hetkellä muodostuvia yhdyssanoja. Tosin tutkimuksen valmistumisen (2015) ja vuosien 2000–2009 välillä periaatteellisia muutoksia ei todennäköisesti ole ehtinyt tapahtua kovin paljon. Tutkimuksen tuloksista voidaankin tehdä myös nykyhetkeen ulottuvia päätelmiä. Selvää on, että tällä hetkellä kieleen syntyvien uusien yhdyssanojen rakenne on todennäköisesti lähempänä tässä kuin vanhemmissa tutkimuksissa kuvattua. Tämä tutkimus on myös ensimmäinen, jossa käytetään nimenomaan uutta yhdyssanastoa sisältävää aineistoa.

Erilaisten aineistojen saatavuus, käytettävyys ja luotettavuus huomioon ottaen tutkimuksen aineistonvalinta on hyvin perusteltu. Tutkimusaineistolla on kuitenkin omat rajoituksensa. Koska tutkimusaineiston pohjana olevaan NSST:aan ei kerätä järjestelmällisesti ortografisia sanaliittoja, tutkimusaineistoa ei ollut mahdollista rajata yhdyssanan teoreettisen määritelmän mukaisesti. Suomen yhdyssanojen oikeinkirjoitussäännöt vastaavat kuitenkin hyvin teoreettista yhdyssanaa (lähinnä taipumattomia ja vajaasti taipuvia sanayhtymiä koskevia poikkeuksia lukuun ottamatta; ks. tarkemmin luku 2.2.1). Oikeinkirjoitusnormit tuntevalle kirjoittajalle

ne ovat myös keino osoittaa kirjoittajan omaa käsitystä sanayhtymän merkityksen tiivyydestä. Koska esimerkiksi varsinkin muilla sijamuodoilla kuin nominatiivilla alkavat sanayhtymät voidaan usein kirjoittaa suomen kielessä sekä yhteen että erikseen merkityksen juuri muuttumatta (esim. *työssäjaksaminen* [TT] ~ *työssä jaksaminen*, *aikaavievä* ~ *aikaa vievä*), tutkimuksen tuloksia ei kuitenkaan voida pitää yksiselitteisen luotettavina kaikkien teoreettisten yhdyssanojen osalta. Ortografisten yhdyssanojen näkökulmasta tulokset sen sijaan ovat luotettavia. (Teoreettisen ja ortografisen yhdyssanan määritelmistä ks. luku 2.2.1.)

Tutkimuksen tuloksia arvioitaessa on pidettävä mielessä myös tutkimusaineiston pohjana olevan NSST:n ensisijainen tarkoitus: siihen kerätään sanastoa ensisijaisesti sanakirjatyön ja kielenhuollon tarpeisiin. NSST sisältää paitsi täysin uutta, myös jonkin verran uudella tavalla käytettyä vanhaa sanastoa. Pelkästään uudessa merkityksessä käytetyt vanhat sanat eivät riitä osoitukseksi rakenteen produktiivisuudesta nykyisessä sananmuodostuksessa. Tämän vuoksi olen kiinnittänyt tässä tutkimuksessa varsinkin harvinaisimpien rakenteiden kohdalla huomiota siihen, kuinka suuri osa rakennetta noudattavista tapauksista on pelkästään semanttisia uudennoksia. Yleisten rakenteiden joukossa semanttisten uudennosten osuus aineistossa on niin pieni, että se ei pääse vääristämään tutkimuksen tuloksia.

Yksi tutkimukseni tarkoituksista on ollut testata uutta metodia ja arvioida sen sopivuutta yhdyssanarakenteiden produktiivisuuden tarkasteluun. Tämä tutkimus on tuonut fennistiikkaan uudenlaisen tavan tarkastella (sana)rakenteiden produktiivisuutta. Uudenlaisen metodin luomiseen on ajanut käytännön pakko: produktiivisuuden tutkimuksessa aiemmin käytetyt menetelmät eivät sellaisenaan sovellu suomen kielen yhdyssanarakenteiden tutkimukseen. Käytetty metodi on yksinkertainen mutta ainakin tämän tutkimuksen valossa toimiva. Metodi yhdistää nimenomaan uudessa sanastossa havaittujen rakenteiden tyyppifrekvenssien kvantitatiivisen analyysin sekä tilastoissa tavalla tai toisella huomiota herättävien rakenteiden kvalitatiivisen tarkastelun (ks. tarkemmin luku 3.3.4). Nähdäkseni nämä lähestymistavat tukevat hyvin toisiaan.

Jo pelkän kvantitatiivisen analyysin avulla voidaan havaita paljon uutta tuottavat produktiiviset rakenteet. Samalla kvantitatiivinen analyysi on hyvä työkalu vain vähän uutta tuottavien ja siksi tarkempaa kvalitatiivista analyysia vaativien rakenteiden löytämiseen. Koska yksittäiset uudet sanat voivat syntyä myös muuten kuin rakenteen produktiivisuuden motivoimina, uuden sanan ilmaantuminen ei aina ole suora osoitus rakenteen produktiivisuudesta (ks. myös Bauer 2001: 56). Kvalitatiivisen tarkastelun käyttöä myös muiden tilastollisten produktiivisuudentutkimusmenetelmien tukena onkin pidetty ehdottoman tärkeänä (esim. Plag 1999: 111–115; Bauer 2001: 150–151). Vähän uutta tuottavien rakenteiden tutkimuksessa kvalitatiivinen tarkastelu on yhtä tärkeää kuin tilastollisesta tarkastelusta saatujen tulosten asiantunteva tulkinta.

Kvalitatiivisen tarkastelun ja useiden tilastollisten menetelmien yhdistäminen parantaa tutkimuksen tulosten luotettavuutta. Erilaisten menetelmien rinnakkainen käyttö myös mahdollistaa sekä aihetta koskevan laajan yleiskuvan että

yksittäisiä muuttujia koskevan yksityiskohtaisen tiedon tavoittamisen. Kvantitatiivisen ja kvalitatiivisen tarkastelun yhdistämisen lisäksi olen tässä tutkimuksessa käyttänyt useita erilaisia tilastollisia menetelmiä. Kaikki menetelmät on valittu siten, että kulloinkin käsiteltävänä olevaan tutkimuskysymykseen voidaan saada mahdollisimman luotettava, havainnollinen ja tarkka vastaus. Erilaisten tilastomenetelmien yhdistelyn ansiosta tutkimuksessa on pystytty kuvaamaan myös muuttujien välillä olevia keskinäisiä suhteita.

8.3 Tulosten käyttö

Tämä tutkimus on luonteeltaan kielen rakennetta kartoittavaa perustutkimusta. Koska tutkimus on ensimmäinen suomen kielen yhdyssanarakenteiden produktiivisuutta käytännöllisen suosion näkökulmasta tarkasteleva työ, sen tarkoitus on ensisijaisesti luoda yleiskuva aiheesta. Samalla tutkimus avaa useita näkökulmia jatkotutkimukselle. Tutkimuksen tuloksia voidaan soveltaa varsinkin opetuksessa ja kieliteknologiassa. Tässä luvussa käsitelen ensin tutkimuksen esiin nostamia jatkotutkimusaiheita (8.3.1) ja sen jälkeen tutkimuksen tulosten käytännön sovellusmahdollisuuksia (8.3.2).

8.3.1 Jatkotutkimus

Produktiivisuuden kuvaus on tärkeä osa kielen morfologian synkronista kokonaiskuvaa, mutta muutokset rakenteiden produktiivisuudessa tulisi huomioida myös kielten diakronisissa kuvauksissa (esim. Bauer 2001: 9). Yleisesti kielen muutoksessa havaittu tendenssi on, että jo entuudestaan suositut rakennepiirteet kasvattavat suosiotaan, sillä kielenkäyttäjien on helppo ottaa niistä mallia uusille muodosteille. Diakronisia muutoksia ei kuitenkaan voida ennustaa pelkästään frekvenssien perusteella (esim. Bauer 2001: 6–8). Yksi keskeinen rakenteiden produktiivisuuteen liittyvä tutkimuskohde onkin rakenteiden suosion syiden tutkimus.

Sananmuodostuksessa käytettävän rakenteen suosio on seurausta siitä, että uutta sanaa luodessaan ihmiset valitsevat mieluummin yhden kuin toisen rakenteen. Jotta sananmuodostuksen tutkimuksessa pystyttäisiin sanomaan, miksi jokin rakenne on toista suosittumpi, olisi päästävä käsiksi niihin (pääasiassa tiedostamattomiin) kielellisiin ja kielenulkoisiin syihin, joiden perusteella kielenkäyttäjät tekevät valintansa uusia sanoja muodostaessaan.¹⁰⁶ Kielenkäyttäjien valintoihin voivat vaikuttaa esimerkiksi tässä tutkimuksessa esiin nostetut käytettävyys, jo käytössä

106 Todellisessa kielenkäytössä eri syyt jonkin tietyn rakenteen valintaan voivat vaikuttaa kaoottisesti yhteen kietoutuneilta ilmiöiltä (erilaisista produktiivisuuteen vaikuttavista seikoista ks. tarkemmin luku 3.2). Tämän vuoksi esimerkiksi neurolingvistiikan alan kokeellisissa tutkimuksissa ärsykeissä muunnellaan systemaattisesti yleensä vain yhtä tai kahta tekijää ja samalla tallennetaan muutosten vaikutukset reaktioihin (Brattico 2013: 103). Toisin kuin tiukasti kontrolloidussa kokeellisessa tutkimuksessa, todellisen kielenkäytön ilmiöitä tarkastelevassa tutkimuksessa ei välttämättä pystytä osoittamaan, mitä tekijöitä ihmisten kieleen liittyvät reaktiot heijastavat (mts. 107).

olevien sanojen tarjoamat mallit, kielenhuollon ohjeet, rakenteeseen liitetty tyyli-arvo sekä muoti. Eri rakenteiden suosio voi olla yhteydessä myös sanojen muodostusmotivaatioon kuten laina- tai kotoperäisyyteen. Merkitystä on myös sosiolingvistiksi seikoilla sekä kulloinkin käytössä olevalla kielimuodolla. Tässä tutkimuksessa ei ole voitu tarkastella kovin syvällisesti kaikkia suomen kielen yhdyssanarakenteiden produktiivisuuteen liittyviä syitä. Tämä tutkimus tarjoaa kuitenkin paljon yhdyssanarakenteiden suosioon liittyvää pohjatietoa suosion syistä kiinnostuneelle tutkijalle.

Osa tutkimusaineiston yhdyssanoista on jo vakiintunut kieleen, osa taas on jäänyt päiväperhoiksi. Koska leksikkoon vakiintuneita yhdyssanoja käytetään tilapäismuodosteita useammin, ne ovat tilapäismuodosteita todennäköisempi malli uudismuodosteille. Lisäksi rakenteeltaan tutuntuntuiset sanat saattavat saada helpommin kieliyhteisön hyväksynnän ja vakiintua helpommin osaksi leksikkoon, jolloin ne vuorostaan tarjoavat todennäköisimmän mallin niiden jälkeen tuleville uusille muodosteille.

Sekä tilapäismuodosteiksi jäävät että myöhemmin leksikkoon vakiintuvat yhdyssanat heijastavat rakenteen synkronista produktiivisuutta yhtä lailla. Produktiivisuuden muutosten kannalta olennaista tietoa olisi sitä vastoin myös se, millaisia rakenteita noudattavat sanat vakiintuvat kieleen. Tämän tutkimuksen tulokset sopivat lähtökohdaksi myös sanaston vakiintumista käsitteleville tutkimuksille. Tuloksia voidaan verrata esimerkiksi uusiin sanakirjoihin otetun yhdyssanaston rakenteeseen. Sanakirja-aineistoa käytettäessä on tosin aina muistettava ottaa huomioon yhdyssanoihin liittyvät rajoitukset (ks. esim. luvut 3.3.1 ja 4.1) ja pohdittava, voidaanko vertailulla todella saada selville, a) millaisia rakenteita noudattava yhdyssanat vakiintuvat kieleen, vai ainoastaan b) millaisia yhdyssanoja sanakirjoihin otetaan.

Vakiintuneiden sanojen rakenne ja produktiivisuus ovat yhdyssanoissa ja johdoksissa keskenään erilaisessa suhteessa. Epäproduktiivisiin johdostyyppisiin kuuluu enemmän leksikaalistunutta sanastoa kuin produktiivisiin tyyppisiin. Looginen selitys tälle on se, että epäproduktiiviseen tyyppiin kuuluvaa johdosta ei ole helppo mieltää johdokseksi vaan se hahmotetaan ennemmin leksikkoon sellaiseen kuuluvaksi kokonaisuudeksi. Yhdyssanat sen sijaan on yleensä helppo hahmottaa alkujaan itsenäisistä leksikaalisista osista koostuviksi (poikkeuksena senkaltaiset erittäin leksikaalistuneet tapaukset kuin *elokuva*, *maailma* ja *yhteiskunta*), ja niissä leksikaalistuneiden tapausten määrän voidaankin ajatella ennemmin edistävän rakenteen produktiivisuutta kuin heikentävän sitä. Tässä mielessä yhdyssanat muistuttavat fraaseja ja syntaktisia konstruktioita, joissa leksikaalistuneet tapaukset näyttävät lisäävän rakenteen produktiivisuutta (esim. Heinonen 2010: 364). Yhdys-

sanarakenteet jäävät siis tässäkin suhteessa morfologian ja syntaksin välimaastoon (ks. myös Mäkisalo 2000a ja 2000b).¹⁰⁷

Kuten olen edellä (luku 8.2) todennut, tämän tutkimuksen tuloksia ei voida yleistää koskemaan kaikkia suomen kielen kielimuotoja tai edes kaikkia kirjoitetun yleiskielen alalajeja. Esimerkiksi Baayenin (1992: 109–110) mukaan produktiivisuus voi olla vahvasti kielimuutosidonnaista, eikä rakennetta voida pitää täysin produktiivisena, ellei se sovi eri kielimuotoihin ja rekistereihin. Tarkemman suomen kielen yhdyssanarakenteita ja niiden produktiivisuutta koskevan kuvan saamiseksi olisikin tärkeää tarkastella yhdyssanarakenteiden asemaa myös muissa kielimuodoissa.

Suomen kielen eri kielimuodot ovat jatkuvassa vuorovaikutuksessa keskenään, ja eri kielimuotoja tutkimalla voidaan saada tietoa myös kielen mahdollisista muutossuunnista. Yhdyssanoja pidetään erityisen keskeisenä ilmaisukeinona erikoiskielissä (ks. esim. Tyysteri 2009). Arkista viestintää puolestaan syytetään usein välinpitämättömyydestä yhdyssanojen oikeinkirjoitussääntöjä kohtaan. Koska yleiskielen sanasto tuntuu toisaalta jatkuvasti lähenevän erikoiskieliä ja toisaalta arkikieltä, pidän yleiskielen muutossuuntien kannalta erityisen tärkeänä yleiskielen ja arkikielen sekä yleiskielen ja erikoiskielen vertailua (oman kielen muiden kielimuotojen lisäksi vaikutteita saadaan tietysti myös vieraista kielistä). Piirteitä, jotka ovat yleisempiä erikoiskielessä tai arkikielessä kuin yleiskielessä, voidaan pitää ns. heikkoina signaaleina tai tulevaisuusmerkkeinä siitä, miten yleiskieli ja siinä hyväksyttävänä pidettävät rakenteet mahdollisesti muuttuvat tulevaisuudessa. Tulevaisuudentutkimuksessa *heikoilla signaaleilla* ja *tulevaisuusmerkeillä* tarkoitetaan orastavia merkkejä asioista, jotka voivat tulevaisuudessa nousta merkittäviksi trendeiksi (tarkemmin esim. Mannermaa 2004; Hiltunen 2008; Kuusi–Hiltunen 2009).¹⁰⁸

Esimerkki jo tapahtumassa olevasta tai tapahtuneesta muutoksesta on yhdysverbien hyväksyttävyyys. Kielenhuolto taisteli pitkään yhdysverbejä vastaan, ja niitä pidettiin sopivina korkeintaan leikkisään arkikieleen. Myöhemmin niitä alettiin puolustaa myös erikoiskieliin sopivina (Vesikansa 1989: 257–258). Vähitellen niihin on alettu suhtautua suopeammin myös yleiskielessä. Toinen esimerkki jo

107 Yhdyssanoja ja johdoksia erottaa myös semanttisen läpinäkyvyyden ja produktiivisuuden suhde. Rakenteen morfologinen ja semanttinen läpinäkyvyys mainitaan usein produktiivisuuden ennakkoehtona. Johtomorfologiassa semanttisella läpinäkyvyydellä tarkoitetaan yleensä sitä, että semanttisesti läpinäkyvän johtimen merkitys on aina sama ja se modifioi kannan merkitystä aina samalla tavalla. Vaikka yhdyssanan osat ovat yleensä helposti tunnistettavia, osien välinen suhde on harvoin tällä tavalla läpinäkyvä. Siksi uusien yhdyssanojen tulkintaan tarvitaan yleensä tietoa kontekstista ja kielenulkoisesta maailmasta. Semanttinen läpinäkyvyys ei siis näytä olevan kovin tärkeää yhdyssanojen produktiivisuuden kannalta. Tästä eivät tosin kaikki tutkijat ole samaa mieltä. Esimerkiksi Booijin (2007: 75–76) mukaan yhdyssanojen produktiivisuus johtuu pitkälti niiden semanttisesta läpinäkyvyydestä: törmätessämme uuteen yhdyssanaan tunnemme jo sen osien merkityksen, joten pääteltäväksi jää vain osien välillä olevan suhde. Toinen Booijin (mp.) mainitsema produktiivisuutta edistävä seikka on yhdyssanojen semanttinen monikäyttöisyys, jonka itse näen pitkälti vastakohtana semanttiselle läpinäkyvyydelle.

108 Heikkojen signaalien lisäksi kielitieteessä olisi nähdäkseni hyvä pohtia myös muiden tulevaisuudentutkimuksen käsitteiden ja menetelmien käytettävyyttä. Tulevaisuudentutkimuksessa mahdollisten tulevaisuuksien hahmottelun tarkoitus on yleensä antaa ihmisille mahdollisuuksia vaikuttaa tulevaisuuteen tai vähintään varautua siihen. Kielitieteessä esimerkiksi kielikontakteihin liittyvien tulevaisuuskenaarioiden hahmottelulla voisi olla tällaista sovellusarvoa.

pitkällä olevasta muutoksesta, johon yleiskielen erikoiskielistyminen on todennäköisesti vaikuttanut, on osiltaan kongruoivien yhdyssanojen katoaminen (toisaalta osiltaan kongruoivat yhdyssanat termistyvät helpommin, toisaalta juuri kongruoimattomuus tekee yhdyssanasta termimäisemmän, ks. tarkemmin luku 5.3.3).

Tämä tutkimus on ensimmäinen osoitus uudenlaisen tutkimusmetodin toimivuudesta yhdyssanarakenteiden tutkimuksessa. Myöhempien tutkimusten tehtäväksi kuitenkin jää metodin luotettavuuden varmistaminen. Yksi parhaista keinoista siihen on selvittää, pitävätkö uudessa sanastossa havaittujen rakenteen tyyppifrekvenssien kvantitatiivisen analyysin sekä varsinkin harvinaisimpien rakenteiden kvalitatiivisen tarkastelun kautta saadut tulokset yhtä muilla tutkimusmetodeilla saatavien tulosten kanssa. Tässä tutkimuksessa tällaista vertailua ei ole voitu tehdä, sillä aiemmin käytettyjen metodien ja yhdyssanatutkimuksen yhteensovittaminen ei toistaiseksi ole mahdollista sopivien aineistojen ja riittävien työkalujen puuttumisen vuoksi (ks. tarkemmin luvut 3.3 ja 4.1). Tulevaisuudessa tämän tutkimuksen tuloksia voidaan kuitenkin hyödyntää kieliteknologiassa (ks. luku 8.3.2) myös yhdyssanojen sisäistä rakennetta varmemmin ja tarkemmin käsittelevien morfologisten jäsentimien kehittämisessä. Uusien jäsentimien myötä ainakin korpuksessa esiintyvien kertamuodosteiden määrään perustuvan metodin (ns. Baayenin metodi) käyttö tulee teknisesti mahdolliseksi myös yhdyssanarakenteiden produktiivisuuden tutkimuksessa.

Tässä tutkimuksessa käytetyn uuden metodin yhtäpitävyyttä aiemmin käytettyjen metodien kanssa voidaan tarkastella myös muunlaisia kielen rakenteita käsittelevissä tutkimuksissa. Metodien soveltaminen ja testaaminen toisenlaisten rakenteiden tutkimuksessa on todennäköisesti myös paras reitti metodin jatkokehittelylle. Sovelluksissa on otettava huomioon erityisesti kaksi seikkaa: 1) Yhdyssanojen muodostus on suomen kielessä erittäin produktiivinen sananmuodostuskeino. Metodi ei välttämättä sovi ainakaan sellaisenaan vähemmän produktiivisten rakenteiden tutkimukseen. 2) Yhdyssanan rakennepiirteille on ominaista, että niistä monet ovat keskenään vaihtoehtoisia. Erilaisilla rakenteilla voidaan siis ilmaista samaa asiaa. Tässä tutkimuksessa parhaat tulokset onkin nähdäkseni saatu nimenomaan keskenään vaihtoehtoisten rakenteiden tarkastelusta. Sen sijaan uusien yhdyssanojen sanaluokkaa koskevien tulosten tulkinta on ollut haastavampaa, koska kyse ei ole keskenään kilpailevista rakenteista. (Ks. myös luku 3.3.4.) Myös johdostyyppit muodostavat ryhmiä, joihin kuuluvat tyyppit voidaan nähdä keskenään kilpailevina (esim. erilaiset teonnimet). Sovellettuna tässä tutkimuksessa käytetty metodi sopii todennäköisesti myös johdosten tutkimukseen, kunhan uudissanastossa havaitut tyyppifrekvenssit ovat riittävän suuria (tosin johdosten tutkimuksessa tyyppiin sopivien kantojen määrä on otettava tarkemmin huomioon kuin yhdyssanatutkimuksessa).

8.3.2 Sovellusmahdollisuudet

Tämän tutkimuksen tulosten kaksi tärkeintä sovellusalaa ovat kieliteknologia ja opetus. Kieliteknologiassa tämän tutkimuksen tuloksia voidaan käyttää apuna kieltä käsittelevien tietokoneohjelmien kuten morfologisten jäsentimien ja

oikolukuohjelmien suunnittelussa. Luonnollisen kielen automaattinen prosessointi perustuu pitkälti todennäköisyyksiin, joista tämä tutkimus tarjoaa runsaasti tietoa. Nykyiset morfologiset jäsentimet osaavat pilkkoa oikein osiinsa useimmat vakiintuneet yhdyssanat. Prosessoitavat aineistot ovat kuitenkin yleensä ennen käsittelemättömiä ja sisältävät myös uudissanoja. Sanaston uusiutumisen vuoksi kaikkea kielen sanastoa ei voida koskaan listata, mutta erityisesti kaikkein marginaalisimpien rakenteiden kohdalla olisi tärkeää pystyä päättämään, sopiiko automaattisen prosessoinnin tarpeisiin paremmin yksittäistapausten listaus vai kyseistä rakennetta kuvaava sääntö.

Täysin korrektiin suomen kielen koneellinen tuottaminen tai esimerkiksi oikolukeminen ei ole mahdollista ilman ajan tasalla olevaa tietoa kielessä produktiivista yhdyssanarakenteista. Suomen kielen varsin eripituiset sanat ja rikas morfologia ovat erityinen haaste yhdyssanojen automaattiselle segmentoinnille ja sisäisen rakenteen tunnistamiselle, eikä kaikkia harvinaisia rakenteita ei ole välttämättä otettu nykyisten jäsentimien suunnittelussa edes huomioon (esim. Lindén–Pirinen 2009). Yksi luonnollisen kielen automaattisen prosessoinnin keskeisistä ongelmista onkin se, että varsinkin harvinaisten rakenteiden analysointi on usein epävarmaa (tarkemmin esim. Evert–Lüdeling 2001). Tämä tutkimus tarjoaa vähintään suuntaa-antavan kehikon sille, mitkä alkuosan muodot ja sanaluokat morfologisille jäsentimille on opetettava, jotta ne selviytyvät suomen kielen yhdyssanojen rakenteen analysoinnista.

Yksittäisten rakennepiirteiden toteutumistodennäköisyyksien lisäksi tämä tutkimus tarjoaa tietoa myös rakennepiirteiden keskinäisistä suhteista (ks. varsinkin luku 7). Luokittavia puita on käytetty tässä tutkimuksessa ensisijaisesti tutkimusaineiston kuvailuun ja muuttujien välillä olevien yhteyksien etsimiseen. Jatkotutkimuksessa kannattaa kuitenkin kartoittaa, kuinka hyvin tämän tutkimuksen aineiston perusteella kasvatetut puut ennustavat muiden samankaltaisten aineistojen jakaumia ja voiko näitä puita hyödyntää muiden aineistojen käsittelyssä. Tutkimuksessa käytetty aineistotietokanta voidaan helposti muuntaa muotoon, jossa sitä voidaan käyttää, kun automaattisen prosessoinnin työkaluille opetetaan niiden prosessoinnissa tarvitsemia tietoja. Tätä mahdollisuutta on ryhdytty jo kartoittamaan Filip Ginterin ja Veronika Laippalan projektissa, jonka tarkoitus on tuottaa koko suomenkieliseen internetiin perustuva tutkijoille vapaasti saatavilla oleva morfosyntaktisesti koodattu kieliaineisto (projektista tarkemmin ks. Laippala–Ginter 2014).

Kuten olen jo luvuissa 8.1 ja 8.2 maininnut, tämän tutkimuksen tulokset on syytä ottaa huomioon tulevia kielioppeja ja sanastonkuvauksia laadittaessa. On myös itsestään selvää, että tulevien suomen kielen asiantuntijoiden on syytä tuntea uusien suomen sananmuodostusta koskeva tieto. Tutkimuksen tulokset ovat helposti yksinkertaistettavissa myös käytännön kielitaitoon tähtäävän opetuksen ja oppikirjojen tarpeisiin. Nähdäkseni tutkimuksen tuomasta uudesta tiedosta voidaan hyötyä erityisesti akateemisissa S2-opetuksessa, mutta oppijoiden lähtötason ja resurssit huomioivia sovelluksia voidaan kehittää myös muuhun opetukseen. Yksi perusopetuksen ja lukion opetussuunnitelmien uudistustyön keskeisistä periaatteista on

kielitietoisuuden lisääminen (kielitietoisuudesta ja opetussuunnitelmista ks. Harmanen 2013), jonka kannalta myös sananmuodostusta koskevan ymmärryksen parantaminen on tärkeää.

S2-opetuksessa yhdyssanatiedon opettaminen jo opintojen alkuvaiheessa tukee sekä uusien sanojen ymmärtämistä että omaa ilmaisuvoimaa. Vaikka suurin osa suomen kieleen muodostettavista uusista sanoista on yhdyssanoja, sananmuodostuksen opetuksessa on perinteisesti keskitytty enemmän johdoksiin kuin yhdyssanoihin. Yhdyssanat tulevatkin S2-oppijoille tutuiksi lähinnä leksikaalistuneiden muodostusten merkityksen opetteluun kautta. Todellisessa kielenkäytössä kuitenkin jopa yli puolet suomen kielen yhdyssanoista saattaa olla leksikaalistumattomia tilapäismuodosteita (Mäkisalo 2000a: 14). Tämän vuoksi olisikin tärkeää, että yhdyssanat huomioitaisiin opetuksessa myös sananmuodostuksen näkökulmasta. Itselle tuntemattomien sanojen ymmärtämisen kannalta oleellista olisi varsinkin taito tunnistaa yhdyssanan osat, mihin taas tarvitaan tietoa esimerkiksi yhdyssanan alkuosan mahdollisista muodoista. Tämä taito voisi samalla hälventää pelkoa suomen kielen pitkiä sanoja kohtaan. Tämä tutkimus on myös osoittanut, että itse asiassa tyypilliset suomen kielen yhdyssanat eivät ole kovin pitkiä (ks. luku 5.1).

Yhdyssanojen muodostuksen opettaminen olisi myös hyvä keino tukea oppijoiden omaa suomenkielistä ilmaisuvoimaa jo opintojen alkuvaiheessa, sillä myös itse uusia sanoja muodostaessaan oppijat muodostavat yleensä enemmän yhdyssanoja kuin johdoksia (esim. Broeder–Extra–van Hout 1996). Alkuvaiheen S2-opetuksen kannalta tutkimuksen tärkeimmät tulokset ovat prototyyppisten tapausten ylivalta uusien yhdyssanojen joukossa sekä se, että monia yhdysosamuotoja käytetään vain yksittäisistä sanoista – ja silloinkin yleensä nominatiivin kanssa vaihdellen. Ymmärrettävien uusien yhdyssanojen muodostamiseen riittää siis yleensä yksi rakenne. Opintojen edistyessä tutuiksi tulevat myös harvinaisemmat rakenteet, mutta epäproduktiivisten rakenteiden opettelu on omien sanojen muodostuksen kannalta todennäköisesti tarpeellista edistyneillekin oppijoille. Itselle vieraiden yhdyssanojen osien tunnistamisessa apua voi olla esimerkiksi yhdyssanassa poikkeuksellisissa muodoissa esiintyvien sanojen (lyhyeksi jäävistä) luetteloista, jollaisia käytetään myös muiden kielen epäsäännöllisyyksien opiskelussa (esim. *neljä* > *neli* [~ *neljä*], *sika* > *sika* - ~ *siko*-). Tulevaisuudessa tällaisia listoja pystytään ehkä tuottamaan automaattisesti. Automaattisen prosessoinnin avulla voidaan tulevaisuudessa todennäköisesti toteuttaa myös erityisesti S2-oppijoita hyödyttävä mahdollisuus erilaisten yhdysosamuotojen käyttöön sanakirjahauissa (esim. hakusanalla *hevos* löytäisi selityksen sanalle *hevon*).

Varsinkin äidinkielisille puhujille suunnattu yhdyssanoja koskeva opetus on perinteisesti keskittynyt ensisijaisesti oikeinkirjoitussääntöihin, mistä sitä on myös kritisoitu (esim. Niinimäki 1991; Ylikulju 2004). Esimerkiksi Niinimäen (1991) mukaan yhdyssanojen oikeinkirjoituksen oppimista voitaisiin parhaiten tukea vahvistamalla opiskelijoiden yhdyssanatajua. Yksi keino niin yhdyssanatajun kuin kielitietoisuudenkin lisäämiseen on kielessä käytettyjen sananmuodostuskeinojen käsittely opetuksessa. Sananmuodostuksen kautta voidaan korostaa esimerkiksi

yhdyssanojen tarjoamia ilmaisumahdollisuuksia: ne ovat helppo keino muodostaa uusia sanoja ja valinnaisissa tapauksissa kirjoittaja voi yhteen ja erikseen kirjoittamisella osoittaa omaa käsitystään yhtymän merkityksen kiteytyneisyydestä. Sanastossa esiintyvien mutta epäproduktiiviseksi muuttuneiden tai muuttumassa olevien rakenteiden käsittelyllä voidaan lisätä ymmärrystä kielen muuttumisesta. Oikeinkirjoituksen lisäksi kielenhuollossa ja opetuksessa on usein kiinnitetty huomiota nominatiivin ja genetiivin vaihteluun yhdyssanan alkuosassa, ja varsinkin genetiivialkuisen yhdyssanarakenteen produktiivisuudessa tapahtuneet ja tapahtuvat muutokset on syytä huomioida sekä opetuksessa että kielenhuollossa.

8.4 Päätössanat

Tämän tutkimuksen tarkoitus on ollut selvittää, mitkä suomen kielessä mahdollisista morfosyntaktisista yhdyssanarakenteista ovat produktiivisia kielenkäytössä. Tutkimus tarjoaa ensimmäistä kertaa tilastollista tietoa siitä, kuinka ahkerasti erilaisia kielen suomia mahdollisuuksia käytetään uusien yhdyssanojen muodostuksessa. Kaiken kaikkiaan tämä tutkimus osoittaa, että suomen kielen uudet yhdyssanat ovat rakenteeltaan huomattavasti homogeenisempi joukko, kuin aiempia kielenkuvauksia lukemalla voisi olettaa. Koska kielenkuvasten tehtävä on yleensä kuvata kaikki se, mikä kielessä on mahdollista, todellisessa kielenkäytössä harvinaiset rakenteet saavat niissä usein suhteettoman paljon huomiota.

Rakenteiden suosiota eli käytännöllistä produktiivisuutta kartoittava tutkimus täydentää kielioppien ja sanastonkuvausten antamaa kuvaa kielestä. Täydentämisen lisäksi tämä tutkimus on tuonut esiin myös sellaisia suomen kielen rakenteeseen liittyviä seikkoja, jotka on hyvä ottaa huomioon tulevia kielenkuvauksia laadittaessa. Suomen kielen yhdyssanojen rakennetta koskeva tieto on fennistien lisäksi tärkeää myös typologeille. Uuden tiedon tuottamisen lisäksi tutkimuksessa esitelty ja testattu uusi metodi edistää produktiivisuuden tutkimusta myös metodologisesti.

Kielenkäyttäjät voivat suomen kielessä valita yhdyssanan rakenteen varsin vapaasti. Tämän tutkimuksen mukaan prototyyppisin ja yksinkertaisin rakenne riittää yleensä kielenkäyttäjille yhdyssanoihin tiivistyvän informaation tarkoitukseenmukaiseen pakkaamiseen. Uusi suomen kielen yhdyssana on todennäköisimmin kahdesta substantiivista yhdistämällä muodostettu substantiivi, jonka alkuosa on nominatiivissa eikä kongruoi jälkiosan kanssa. Tällainen rakenne on periaatteessa mahdollinen millaisessa tahansa merkitystehtävässä, joten se on aina mahdollinen vaihtoehto jollekin viittauskohdettaan tarkemmin kuvaavalle rakenteelle. Nominatiivialkuinen rakenne jättää merkitystulkinnan hyvin vapaaksi, mutta maailmantiedon ja viimeistään kontekstin avulla ymmärretään sellaisetkin sanat kuin *allergianäytäntö*, *helikopteriäiti*, *korvakynttilä*, *pesäkieli* ja *ötökkätarjonta*.

Prototyyppisten tapausten lisäksi myös monenlaiset muut rakenteet ovat kielopin tasolla mahdollisia. Käytännössä kaikki niistä eivät kuitenkaan juuri tuota

uutta sanastoa. Osa aiemmassa tutkimuksessa produktiivisiksi mainituista yhdysanarakenteista on uudessa yhdysanastossa niin harvinaisia, että niitä ei voida enää pitää produktiivisina. Esimerkiksi osiltaan kongruoivia yhdysanoja syntyy hyvin vähän (esim. *mustasuola* : *mustansuolan* ~ *mustasuolan*). Ennako-oletusta harvinaisempia ovat myös genetiivialkuiset yhdysanat (esim. *koodinpätkä*, *sisälöntuottaja*). Prefiksinkaltaisella alkuosalla alkavat yhdysanat (esim. *lähimummo*, *tasapalkka*) puolestaan ovat huomattavasti yleisempiä, kuin sanastonkuvauksia lukemalla voisi olettaa. Erilaisista ainoastaan yhdysanan alkuosana esiintyvistä yhdysosamuodoista edelleen mahdollisia ovat kaikki sanastonkuvauksissamme mainitut. Osaa yhdysosamuodoista käytetään kuitenkin vain yksittäisistä sanoista, eikä niitä siksi voida pitää produktiivisina rakenteellisella tasolla.

Odotuksenmukaisesti valtaosa suomen kielen uusista yhdysanoista on substantiiveja. Uutta yhdysanastoa syntyy kuitenkin jonkin verran myös muihin sanaluokkiin (esim. *moniammatillinen*, *täsmäsiivota*, *yhteistoiminnallisesti*). Muihin sanaluokkiin kuuluvissa yhdysanoissa alkuosa on muu kuin nominatiivimuotoinen substantiivi useammin kuin yhdysubstantiiveissa. Epätavalliset piirteet siis kertyvät jossain määrin samoihin yhdysanoihin. Epätavallisista piirteistä ainoastaan moniosaisuus on nimenomaan yhdysubstantiivien ominaisuus, sillä muihin sanaluokkiin kuuluvissa sanoissa on harvoin enemmän kuin kaksi leksikaalista osaa. Yhdysubstantiiveissa mahdollisia ovat *kehitysmaasairaanhoidajakoulutuksen* ja *tuplapuoliperävaunuyhdistelmän* kaltaiset hyvin pitkät muodosteet, mutta yhdysubstantiiveissakaan on harvoin enemmän kuin neljä leksikaalista osaa.

Tämä tutkimus tarjoaa paljon tietoa siitä, millaista yhdysanojen muodostus on ollut suomen kielessä 2000-luvun alussa. Aika näyttää, miten suomen kielen systeemi muuttuu tulevaisuudessa yhdysanojen osalta. Tämän tutkimuksen perusteella silmällä pidettäviä tendenssejä ovat toisaalta yhdysanaston muuttuminen rakenteeltaan prototyypiseen suuntaan ja toisaalta varsinkin prefiksinkaltaisten alkuosien synnyttämä prototyypin uusi murtumakohta.

Lähteet

- Agresti, Alan 2002: *Categorical data analysis*. Second edition. Wiley Series in probability and statistics. Wiley-Interscience, New York.
- Airila, Martti 1938: Suomen kielen yhdysadjektiivit. *Vir.* 42 (3), 357–363.
- Akama, Hiroyuki 2003: *Probabilistic language processing in the form of a decision tree: Usage of the French impersonal subject pronoun "on" (or "l'on")*. Verkkoartikkeli osoitteessa <http://www.dp.hum.titech.ac.jp/~akama/on.pdf> (viitattu 27.10.2014). Tokyo Institute of Technology, Tokyo.
- Alho, Irja – Kauppinen, Anneli 2008: *Käyttökielioppi*. SKS, Helsinki.
- Allen, Margaret Reece 1978: *Morphological investigations*. Diss. University of Connecticut, Storrs, CT.
- Anderson, Stephen R. 1982: Where's morphology? *Linguistic Inquiry* 13 (4), 571–612.
- Argus, Reili – Johansen Ijä, Johanna – Laalo, Klaus 2014: Liitnimisõnade omandamine eesti, soome ja saami keeles: ühist ja erinevat. *Keel ja kirjandus* (8–9), 648–669.
- Argus, Reili – Kazakovskaya, Victoria 2013: Acquisition of compounds in Estonian and Russian: Frequency, productivity, transparency and simplicity effect. *Eesti Rakenduslingvistika Ühingu aastaraamat* 9, 23–42.
- Arjava, Hellevi 2005: *Alpiini ja Otteljaana. Kangasniemeläisten etunimet 1684–1899*. Suomi 190. SKS, Helsinki.
- Aronoff, Mark 1976: *Word formation in generative grammar*. Linguistic Inquiry Monograph One. The MIT Press, Cambridge–Massachusetts.
- 1983: Potential words, actual words, productivity and frequency. *Proceedings of the 13th International Congress of Linguists*, 163–171. Oslo University Press, Oslo.
- Aronoff, Mark – Schvaneveldt, Roger 1978: Testing morphological productivity. *Annals of the New York Academy of Sciences* 318 (1), 106–114.
- Baayen, R. Harald 1989: *A corpus-based approach to morphological productivity: statistical analysis and psycholinguistic interpretation*. Diss. Centrum voor Wiskunde en informatica, Free University of Amsterdam.
- 1992: Quantitative aspects of morphological productivity. *Yearbook of morphology* 1991, 109–149. Kluwer, Dordrecht–Boston–London.
- 1993: On frequency, transparency and productivity. *Yearbook of morphology* 1992, 181–208. Kluwer, Dordrecht–Boston–London.
- Baayen, R. Harald – Lieber, Rochelle 1991: Productivity and English derivation: A corpus-based study. *Linguistics* 29, 801–843.
- Baayen, R. Harald – Renouf, Antoinette 1996: Chronicling the times: Productive lexical innovations in an English newspaper. *Language* 72 (1), 69–96.
- Baayen, R. Harald – Tweedie, Fiona J. – Schreuder, Robert 2002: The subject as a simple random effect fallacy: subject variability and morphological family effects in the mental lexicon. *Brain and Language* 81 (1), 55–65.
- Baker, Mark 1987: *Incorporation: A theory of grammatical function changing*. University of Chicago Press, Chicago.
- Barðdal, Jóhanna 2008: *Productivity. Evidence from case and argument structure in Icelandic*. John Benjamins, Amsterdam–Philadelphia.
- Bauer, Laurie 1978: *The grammar of nominal compounding: with special reference to Danish, English and French*. Odense UP, Odense.
- 1988: *Introducing linguistic morphology*. Edinburgh University Press, Edinburgh.
- 1998: When is a sequence of two nouns a compound in English? *English Language and Linguistics (EnLL)* 2 (1), 65–86.
- 2001: *Morphological productivity*. Cambridge University Press, Cambridge.
- 2005: Productivity: Theories. *Handbook of Word-Formation*, 315–334. Toim. Pavol Štekauer – Rochelle Lieber. Studies in Natural Language and Linguistic Theory 64. Springer Netherlands, Dordrecht.
- 2009a: Danish. *OHC*, 400–416.

- Bauer, Laurie 2009b: Typology of compounds. *OHC*, 343–356.
- Berg, D. L. 1991: *A User's Guide to the OED*. Oxford University Press, Oxford.
- Berko, Jean 1958: The child's learning of English morphology. *Word* 14, 150–177.
- Bertram, Raymond – Hyönä, Jukka 2003: The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language* 48, 615–634. Saatavissa myös⁸⁹: http://users.utu.fi/hyona/Bertram&Hy%C3%B6n%C3%A4_JML2003.pdf.
- 2013: The role of hyphens at the constituent boundary in compound word identification: Facilitative for long, detrimental for short compound words. *Experimental Psychology* 60 (3), 157–63. Saatavissa myös: http://users.utu.fi/hyona/Bertram&Hy%C3%B6n%C3%A4_EP_2013.pdf.
- Bertram, R. – Laine, M. – Karvinen, K. 1999: The interplay of word formation type, affixal homonymy and processing in lexical processing: Evidence from a morphologically rich language. *Journal of Psycholinguistic Research* 28 (3), 213–266.
- Bertram, R. – Pollatsek, A. – Hyönä, J. 2004: Morphological parsing and the use of segmentation cues in reading Finnish compounds. *Journal of Memory and Language* 51, 325–345. Saatavissa myös: <http://users.utu.fi/hyona/Bertam%20et%20al.%202004.pdf>.
- Birdlife: periaatteet = Lintujen nimeämisen periaatteet. Verkkosivu osoitteessa <http://www.birdlife.fi/lintuharrastus/nimisto/periaatteet.shtml> (viitattu 16.1.2014). Birdlife Suomi ry, Helsinki.
- Bisetto, Antonietta – Scalise, Sergio 2005: Classification of compounds. *Lingue e Linguaggio* IV (2), 319–332. Saatavissa myös www.morbocomp.sslmit.unibo.it/download/classification_of_compounds.pdf.
- 2009: The Classification of compounds. *OHC*, 34–53.
- Bloomfield, Leonard 1933: *Language*. Henry Holt, New York.
- Bolozky, Shmuel 1999: *Measuring productivity in word formation. The case of Israeli Hebrew*. Brill, Leiden.
- Booij, Geert 2007: *The grammar of words. An introduction to linguistic morphology*. Oxford University Press, Oxford.
- 2009: Compounding and construction morphology. *OHC*, 201–216.
- Boudelaa, Sami – Marslen-Wilson, William D. 2011: Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes* 26 (4–6), 624–652.
- Brattico, Pauli 2013: Introspektio, kielen autonomia ja aineistot: pohdintoja kielitieteen menetelmistä. *Vir.* 117 (1), 102–109.
- Brattico, Pauli – Leinonen, Alina – Krause, Christina M. 2007: On the limits of productive word formation: Experimental data from Finnish. *SKY Journal of Linguistics* 20, 109–139. Saatavissa myös: <http://www.linguistics.fi/julkaisut/SKY2007/BRATTICO%20ET%20AL.pdf>.
- Breiman ym. 1984 = Breiman, L. – Friedman, J. H. – Olshen, R. A. – Stone, C. J. 1984: *Classification and regression trees*. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA.
- Broeder, Peter – Extra, Guus – van Hout, Roeland 1996: Word-formation processes in adult language acquisition: A multiple case study on Turkish and Moroccan learners of Dutch. *Jyväskylä cross-language studies* 17, 15–24. Toim. Kari Sajavaara – Courtney Fairweather. University of Jyväskylä, Jyväskylä.
- Brown, Morton B. – Forsythe, Alan B. 1974: Robust tests for equality of variances. *Journal of the American Statistical Association* 69 (346), 364–367. Saatavissa myös: <http://drsmorey.org/bibtex/upload/Brown:Forsythe:1974.pdf>.
- Bybee, Joan L. 1985: *Morphology. A study of the relation between meaning and form*. Typological studies in language 9. John Benjamins, Amsterdam.
- 2006: From usage to grammar: the mind's response to repetition. *Language* 82 (4), 711–733.

89 Painetuille lähteille on annettu lähdeluettelossa sähköisiä saatavuustietoja vain, jos sähköiset versiot ovat vapaasti saatavissa. Kaikkien mainittujen internetlähteiden saatavuus on tarkistettu 27.10.2014–30.10.2014. Monet lähdeluettelossa mainituista painetuista artikkeleista ovat saatavissa salasanasuojatuilta sivuilta myös sähköisesti (ks. tarkemmin esim. www.nelliportaali.fi ja oman kirjastosi internetsivut).

- Cabré, Teresa M. 1999: *Terminology. Theory, methods, and applications*. Toim. Juan C. Sager. Käant. Janet Ann DeCesaris. John Benjamins, Amsterdam–Philadelphia.
- CD-PS = *CD-perussanakirja*. KKTKJ 94. KKTK, Helsinki 1997.
- Cellard, Jacques – Sommant, Micheline 1979: *500 mots nouveaux définis et expliqués*. Duculot, Paris.
- Chomsky, Noam 1965: *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.
- 1970: Remarks on nominalization. *Readings in English transformational grammar*, 184–221. Toim. Roderick A. Jacobs – Peter S. Rosenbaum. Ginn, Waltham, MA.
- Cichocki, A. – Amari, S. 2010: Families of alpha-, beta- and gamma-divergences: Flexible and robust measures of similarities. *Entropy* 12 (6), 1532–1568.
- Clark, Brady 2005: On stochastic grammar. *Language* 81 (1), 207–217.
- Coseriu, Eugenio 1971: *Sprache. Strukturen und Funktionen. XII Aufsätze*. 2., verbesserte Auflage. Spangenberg KG, Tübingen.
- 1975: *Sprachtheorie und allgemeine Sprachwissenschaft*. Bd. 2. Internationale Bibliothek für allgemeine Linguistik, München.
- Croft, William 1996: What's a head? *Phrase Structure and the Lexicon*, 35–75. Toim. Johan Rooryck – Laurie Zaring. Studies in Natural Language and Linguistic Theory 33. Kluwer, Dordrecht.
- 2000: *Explaining language change: An evolutionary approach*. Longman, London.
- 2003: *Typology and universals*. 2. edition. Cambridge University Press, Cambridge.
- Cross, T. G. 1977: Mothers' speech adjustments: the contributions of selected child listener variables. *Talking to children: language input and acquisition*, 151–188. Toim. C. Snow – C. A. Ferguson. Cambridge University Press, Cambridge.
- Crystal, David 2006: *Language and the Internet*. 2. edition. Cambridge University Press, Cambridge.
- Dąbrowska, Ewa 2008: The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language* 58, 931–951.
- Downing, Pamela 1977: On the creation and use of English compound nouns. *Language* 53 (4), 810–842.
- Eddington, David 2010: A comparison of two tools for analyzing linguistic data: logistic regression and decision trees. *Italian Journal of Linguistics* 22, 265–286. Saatavissa myös: <http://linguistica.sns.it/RdL/22.2/Eddington.pdf>.
- Elomaa, Marjatta 1984: Suomen kielen sähkötekniillisen termistön muutoksia 20 vuoden aikana. *Erikoiskielet ja käännteoria: Vakki-seminaari IV, Vöyri 28.–29.1.1984*, 7–16. Vaasan korkeakoulu, Vaasa.
- Ernestus, Mirjam – Baayen, R. Harald 2003: Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79 (1), 5–38.
- 2004: Analogical effects in regular past tense production in Dutch. *Linguistics* 42 (5), 873–903.
- Eronen, Riitta 1996: Monimuotoiset yhdyssanat. *Kielikello* (4), 12–14
- 2007: *Uudissanat rötösherrasta salarakkaaseen*. Otava, Helsinki.
- 2010: *Ei- vai epätoivottu tapaus?* *Kielikello* (4), 40.
- 2012: *Minäkertoja vai minä-kertoja*. *Kielikello* (3), 40.
- Eronen, Riitta – Maamies, Sari – Räikkälä, Anneli 1996: Yhdyssanat. *Kielikello* (4), 5–12.
- Evert, Stefan – Lüdeling, Anke 2001: Measuring morphological productivity: Is automatic preprocessing sufficient. *Proceedings of the Corpus Linguistics 2001 conference*, 167–175. Lancaster University. Saatavissa myös: <http://www.stefan-evert.de/PUB/EvertLuedeling2001.pdf>.
- Feise, Ronald J. 2002: Do multiple outcome measures require p-value adjustment? *BMC Medical Research Methodology* 2 (8). Saatavissa myös: <http://www.biomedcentral.com/1471-2288/2/8>.
- Fernández-Domínguez, Jesús 2009: *Productivity in English word-formation: an approach to N+N compounding*. Peter Lang, Bern.
- Fillmore, Charles J. – Kay, Paul – O'Connor, Mary Catherine 1988: Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64 (3), 501–538.
- Fiorentino, Robert – Poeppel, David 2007: Compound words and the structure of the lexicon. *Language and Cognitive Processes* 22 (7), 953–1000. Saatavissa myös: http://www.biolinguistics.uqam.ca/Fiorentino_Poeppel_2007.pdf.
- Fischer, Roswitha 1998: *Lexical change in present-day English. A corpus-based study of the motivation, institutionalization, and productivity of creative neologisms*. Gunter Narr Verlag, Tübingen.

- Fried, Mirjam – Östman, Jan-Ola 2005: Construction grammar: A thumbnail sketch. *Construction grammar in a cross-language perspective*, 11–86. Toim. Mirjam Fried – Jan-Ola Östman. Benjamins, Amsterdam.
- FSNordic.net: Tyhmiä kysymyksiä = Keskustelupuheenvuoro ”Re: Tyhmiä kysymyksiä muusta kuin ilmailusta, tietokoneista tms.” verkko-osoitteessa <http://www.fsnordic.net/discussion/index.php?action=printpage;topic=79167.0> (julkaistu 7.5.2007, poimittu 10.5.2011).
- Gagné, Christina L. 2001: Relation and lexical priming during the interpretation of noun–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27 (1), 236–254.
- 2009: Psycholinguistic perspectives. *OHC*, 255–271.
- Gagné, Christina L. – Shoben, Edward J. 1997: Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23 (1), 71–87.
- Gelman, Andrew – Hill, Jennifer – Yajima, Masanao 2012: Why we (usually) don’t have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness* 5 (2), 189–211.
- Genetz, A. 1881: *Suomen kielen äänne- ja muoto-oppi ynnä runousoppi*. Helsinki.
- Giegerich, Heinz J. 2009: The English compound stress myth. *Word Structure* 2 (1), 1–17.
- Goldberg, Adele 1995: *Constructions. A construction grammar approach to argument structure*. University of Chicago press, Chicago.
- 2006: *Constructions at work: the nature of generalization in language*. Oxford University Press, Oxford.
- Grainger, J. – Ferrand, L. 1994: Phonology and orthography in visual word recognition: Effects of masked homophone primes. *Journal of Memory and Language* 33, 218–233. Saatavissa myös: <http://sites.univ-provence.fr/wlpc/pagesperso/grainger/pub19/Grainger&Ferrand-94.pdf>.
- Grainger ym. 2003 = Grainger, J. – Diependaele, K. – Spinelli, E. – Ferrand, L. – Farioli, F. 2003: Masked repetition and phonological priming within and across modalities. *Journal of Experimental Psychology: Learning, Memory and Cognition* 29 (6), 1256–1269.
- Grandi, Nicola 2009: When morphology ”feeds” syntax: Remarks on noun > adjective conversion in Italian appositive compounds. *Selected Proceedings of the 6th Décembrettes*, 111–124. Toim. Fabio Montermini – Gilles Boyé – Jesse Tseng. Cascadilla Proceedings Project, Somerville, MA. Saatavissa myös: <http://www.lingref.com/cpp/decemb/6/paper2239.pdf>.
- Granger, Sylviane 2002: Computer learner corpora, second language acquisition, and foreign language teaching. *A bird’s eye view of learner corpus research*, 3–33. Toim. Stephanie Petch-Tyson – Joseph Hung – Sylviane Granger. John Benjamins, Amsterdam.
- Greenberg, Joseph H. 1963: Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. *Universals of Human Language*, 73–113. Toim. Joseph H. Greenberg. MIT Press, Cambridge, Mass.
- Gries, Stefan Th. 2009: *Statistics for Linguistics with R: A Practical Introduction*. De Gruyter Mouton, Berlin.
- 2010: Useful corpus-linguistic statistics. *A Mosaic of Corpus Linguistics: Selected Approaches*, 269–291. Toim. Aquilino Sánchez – Moisés Almela. Peter Lang, Frankfurt am Main.
- Grossmann, Maria 2012: Romanian compounds. *Probus* 24 (1), 147–173.
- Grönros, Eija-Riitta 1991: Miten sanakirja syntyy? *Kielikello* (3), 33–34.
- 1994: Yhdysverbit. *Kielikello* (3), 33.
- 2002: Uudissanojen seulonta sanakirjatyössä. *Terminfo* (1), 8–10.
- 2004: Kielitoimiston sanakirja. *Kielikello* (4), 4–8.
- 2006: Kielitoimiston sanakirjasta painettu versio. *Kielikello* (3), 31.
- 2007: Mies ja nainen sanakirjassa. *Kielikello* (4), 13–15
- Haarala, Risto 1981: *Sanastotyön opas*. KKTKJ 16. KKTKJ, Helsinki.
- 1989: Sanat tiedon ja taidon palveluksessa. *Nyky-suomen sanavarat*, 259–275. Toim. Jouko Vesikansa. WSOY, Helsinki.
- Haarala, Risto – Nissinen, Irma 1994: Perussanakirjan uudissanat. *Kielikello* (3), 23–25.
- ten Hacken, Pius 2009: Early generative approaches. *OHC*, 54–77.
- Hakanen, Aimo 1973: *Adjektiivien vastakohtasuhteet suomen kielessä*. SKST 311. SKS, Helsinki.
- Hakulinen, Auli 2006: Ison suomen kielioipin ratkaisuja. *XXXII Kielitieteen päivät Tampereella* 19.–

- 20.5.2005. *Tampere Studies in Language, Translation and Culture. Series B 2*, 335–346. Toim. Anneli Pajunen – Hannu Tommola. Tampere University Press, Tampere.
- Hakulinen, Auli – Karlsson, Fred 1979: *Nyky-suomen lauseoppia*. SKST 350. SKS, Helsinki.
- Hakulinen, Auli – Karlsson, Fred – Vilkkuna, Maria 1996 [1980]: *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus*. Publications No. 6. Department of General Linguistics, University of Helsinki, Helsinki.
- Hakulinen, Auli – Ojanen, Jussi 1976 [1970]: *Kielitieteen ja fonetiikan termistöä*. 2., täydennetty painos. SKST 324. SKS, Helsinki.
- Hakulinen, Lauri 1969: *Suomen sanaston käänntslainoja*. SKST 293. SKS, Helsinki.
- 1979: *Suomen kielen rakenne ja kehitys*. Neljäs, korjattu ja lisätty painos. Otava, Helsinki.
- Halle, Morris 1973: Prolegomena to a theory of word formation. *Linguistic Inquiry* 4 (1), 3–16.
- Harmanen, Minna 2013: Kieli- ja tekstitietoisuutta kouluun! Kielitietoinen koulu ja äidinkielen ja kirjallisuuden opetus. *Kieliverkoston verkkolehti*. Verkkootikkeli osoitteessa <http://www.kieliverkosto.fi/article/kieli-ja-tekstitietoisuutta-kouluun-kielitietoinen-koulu-ja-aidinkielen-ja-kirjallisuuden-opetus/> (viitattu 31.10.2014). Kielikoulutuspolitiikan verkosto, Jyväskylä.
- Havas, Viktória – Rodríguez-Fornells, Antoni – Clahsen, Harald 2012: Brain potentials for derivational morphology: An ERP study of deadjectival nominalizations in Spanish. *Brain and Language* 120 (3), 332–344.
- Hay, Jennifer B. – Baayen, Harald 2005: Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences* 9 (7), 342–348. Saatavissa myös: http://idiom.ucsd.edu/~ackerman/Szuperkursus/Hay_Baayen.pdf.
- Heikkinen, Vesa – Lehtinen, Outi – Lounela, Mikko 2002: Taulukot. Verkkosivu osoitteessa <http://www.kotus.fi/index.phtml?s=2273> (viitattu 14.11.2013). KKK, Helsinki.
- Heinonen, Tarja Riitta 2001: Harmaaturkit herkkusuut – bahuu-vriihit sanakirjassa ja kieliopissa. *Vir*. 105 (4), 625–634.
- 2010: *Kuin-vertaukset*. *Vir*. 114 (3), 348–373.
- 2013: *Idiomien leksikaalinen kuvaus kielenkäytön ja vaihtelun näkökulmasta*. Yleinen kielitiede, Nykykielten laitos, Helsingin yliopisto. Saatavissa myös: https://helda.helsinki.fi/bitstream/handle/10138/37918/heinonen_vaitoskirja.pdf.
- Helasvuo, Marja-Liisa 2014: ”Jotta suomalaiset voisivat puhua enemmän”. Puhetilanteen osallistujat tekstiviestikeskustelussa. *Kieli verkossa. Näkökulmia digitaaliseen vuorovaikutukseen*, 29–49. Toim. Marja-Liisa Helasvuo – Marjut Johansson – Sanna-Kaisa Tanskanen. SKS, Helsinki.
- Heyvaert, Liesbet 2009: Compounding in cognitive Linguistics. *OHC*, 231–254.
- Hiidenmaa, Pirjo 2003: *Suomen kieli – who cares?* Otava, Helsinki.
- Hiltunen, Elina 2008: The future sign and its three dimensions. *Futures* 40 (3), 247–260.
- Holm, S. 1979: A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6, 65–70.
- HS = *Helsingin Sanomat*
- Huumo, Tuomas – Ojutkangas, Krista 2005: Muotikasta kielioppia – jatkumot ja muotit Ison suomen kieliopin selkärankana. *SJ* 47, 222–233.
- Hyönä, Jukka 1995: Pikalukeminen modernin silmänliiketutkimuksen valossa. *Kirjoitetun kielen prosessointi*, 103–110. Toim. Jukka Hyönä – Heikki Lang – Marja Vauras. Oppimistutkimuksen keskus, Turun yliopisto, Turku.
- 2012: The role of visual acuity and segmentation cues in compound word identification. *Frontiers in Psychology* 11 (3), article 188. Verkkootikkeli osoitteessa: <http://journal.frontiersin.org/Journal/10.3389/fpsyg.2012.00188/full> (viitattu 31.10.2014).
- Hyönä, J. – Bertram, R. – Pollatsek, A. 2004: Are long compound words identified serially via their constituents? Evidence from an eye-movement-contingent display change study. *Memory & Cognition* 32 (4), 523–532.
- Hyönä, J. – Niemi, P. 1990: Eye movements during repeated reading of a text. *Acta Psychologica* 73, 259–280. Saatavissa myös: http://users.utu.fi/hyona/Acta_1990.pdf.
- Hyönä, J. – Niemi, P. – Underwood, G. 1989: Reading long words embedded in sentences: Informativeness of word halves affects eye movements. *Journal of Experimental Psychology: Human Perception and Performance* 15 (1), 142–152.
- Hyönä, Jukka – Pollatsek, A. 1998: Reading Finnish compound words: Eye fixations are affected

- by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance* 24 (6), 1612–1627.
- Hyönä, Jukka – Vainio, Seppo 2009: Sentence parsing in a morphologically rich language – Finnish. *Language and Linguistics Compass* (3), 719–733. Saatavissa myös: http://users.utu.fi/hyona/Hy%C3%B6n%C3%A4&Vainio_Compas.pdf.
- Häkkinen, Kaisa 1987: Suomen kielen vanhoista ja uusista yhdysverbeistä. *SJ* 29, 7–29.
- 1990: *Mistä sanat tulevat. Suomalaista etymologiaa*. Tietolipas 117. SKS, Helsinki.
- 2003: Suomettaren pojat. *Suomettaren helmoista. Tutkielmia 1800-luvun suomen kielestä*, 13–43. Toim. Tommi Kurki – Päivi Laine – Arja Lampinen. Kirja-Aurora, Turku.
- 2005: Iso suomen kielioppi osana suomalaisen kieliopin kirjoituksen perinnettä. *Vir.* 109 (4), 554–563.
- Häkkinen, Kaisa – Björkqvist, Pia 1991: Varhaisnykysuomen yhdyssanat. *SJ* 33, 31–55.
- Hämäläinen, K. 1883: *Suomen kielen muoto-oppi*. Helsinki.
- Iikkanen, Leila 1993: *Neologismien muodostustavat ja niiden nykyinen produktiivisuus suomen ja ranskan kielissä*. Pro gradu -tutkielma. Kääntäjänkoulutuslaitos, Turun yliopisto.
- Iisa, Katariina – Oittinen, Hannu – Piehl, Aino 2002: *Kielenhuollon käsikirja*. 5. laajennettu ja muutettu painos; uudistetun ja päivitetyn laitoksen 1. painos. Yrityskirjat, Helsinki.
- Iisa, Katariina – Piehl, Aino 1992: *Virkakielestä kaikkien kieleen*. KKTKJ 67. VAPK-kustannus, Helsinki.
- Ikola, Osmo 1968: Suomen kielioppi ja oikeakielisyysoapas. *Suomen kielen käsikirja*, 103–351. Toim. Osmo Ikola. Weilin + Göös, Helsinki.
- Inhoff, A. W. – Radach, R. – Heller, D. 2000: Complex compounds in German: Interword spaces facilitates segmentation but hinder assignment of meaning. *Journal of Memory and Language* 42 (1), 23–50.
- ISK = *Iso suomen kielioppi*. Hakulinen, Auli – Vilkuna, Maria – Korhonen, Riitta – Koivisto, Vesa – Heinonen, Tarja Riitta – Alho, Irja 2004. SKST 950. SKS, Helsinki. Verkkoversio saatavissa: <http://scripta.kotus.fi/visk/etusivu.php>.
- Itkonen, Erkki 1966: *Kieli ja sen tutkimus*. WSOY, Helsinki.
- Itkonen, Esa 2005: *Analogy as structure and process. Approaches in linguistic, cognitive psychology, and philosophy of science*. J. Benjamins, Philadelphia.
- Itkonen, Terho 2007: *Uusi kielioppi*. 3., tarkistettu painos. Tammi, Helsinki.
- Jaakola, Minna 2004: *Suomen genetiivi*. SKST 995. SKS, Helsinki.
- van Jaarsveld, H. – Rattink, G. 1988: Frequency effects in the processing of lexicalised and novel nominal compounds. *Journal of Psycholinguistic Research* 17 (6), 447–476.
- Jackendoff, Ray 2009: Parallel architecture and conceptual semantics. *OHC*, 105–128.
- Jantunen, Jarmo Harri 2004: *Synonymia ja käännoosuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännoskielen leksikaalisiin erityispiirteisiin*. Joensuun yliopiston humanistisia julkaisuja 35. Joensuun yliopisto, Joensuu.
- Jantunen, Mika 1997: *Havaita, hillitä* -tyyppisten verbien produktiivisuudesta. *Vir.* 101 (3), 434–438.
- Jarema ym. 1999 = Jarema, G. – Bussan, C. – Nikolova, R. – Tsapkini, K. – Libben, B. 1999: Processing compounds: A cross-linguistic study. *Brain and Language* 68 (1–2), 362–369.
- Joki, Leena 2009: Kotuksen sanatieoista. [Henkilökohtainen sähköpostiviesti 25.5.2009]. Kirjoittajan hallussa.
- Jomppanen, Marjatta 2011: Pohjoissaamen ja suomen kielen koloratiivikonstruktio vertailussa. *Lähivertailuja* 21, 106–131. Päätoim. Annekatrin Kaivapalu. Eesti Rakenduslingvistika Ühing, Tallinn.
- de Jong ym. 2002 = de Jong, N. H. – Feldman, L. B. – Schreuder, R. – Pastizzo, P. – Baayen, R. H. 2002: The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects. *Brain and Language* 81 (1–3), 555–567.
- de Jong, N. H. – Schreuder, R. – Baayen R. H. 2000: The morphological family size effect and morphology. *Language and Cognitive Processes* 15 (4), 329–365.
- Jussila, Raimo 1988: Agricolan sanasto ja nykysuomi. *Mikael Agricolan kieli*, 203–228. Toim. Esko Koivusalo. Tietolipas 112. SKS, Helsinki.
- Kanervamäki, Heli 1996: *Yhteen ja erikseen kirjoittamisen normiston kehitys 1900-luvulla*. Pro gradu -tutkielma. Suomalaisen ja yleisen kielitieteen laitos, Turun yliopisto, Turku.
- Kapiainen, Salme 1983: *Nominatiivi ja genetiivi yhdyssubstantiivin määrteisana suomen murteissa*.

- Pro gradu -tutkielma. Suomen kieli, Turun yliopisto.
- Karlsson, Fred 1983: *Suomen kielen äänne- ja muotorakenne*. WSOY, Porvoo–Helsinki–Juva.
- 1994: *Yleinen kielitiede*. Yliopistopaino, Helsinki.
- Karvinen, Marika 1999: *Vieraiden kielten mallin mukaiset yhdysadjektiivit sanomalehti Karjalaisessa*. Pro gradu -tutkielma. Joensuun yliopiston suomen kielen ja kulttuuritieteiden laitos, Joensuun yliopisto.
- Kasik, Reet 1980: Johto-opillisesta produktiivisuudesta ja aktiivisuudesta. *CIFU V:III, Congressus Quintus Internationalis Fenno-Ugristarum: Turku 20.–27. VIII.1980*, 57–63. Toim. Osmo Ikola. Suomen Kielen Seura, Turku.
- Kastovsky, Dieter 2009: Diachronic perspectives. *OHC*, 323–340.
- Kavka, Stanislav 2009: Compounding and idiomatology. *OHC*, 19–33.
- Kay, Paul 1995: Construction grammar. *Handbook of pragmatics. Manual*, 171–177. Toim. Jef Verschueren – Jan-Ola Östman – Jan Blommaert. Benjamins, Amsterdam.
- Keller, Rudi – Nerlich, Brigitte 1994: *On language change. The invisible hand in language*. Routledge, Florence, KY.
- Kenesei, István 2007: Semiwords and affixoids: The territory between word and affix. *Acta Linguistica Hungarica* 54 (3), 263–293. Saatavissa myös: <http://www.nytud.hu/kenesei/publ/semiword.pdf>.
- Kerremans, Daphné – Stegmayr, Susanne – Schmid, Hans-Jörg 2011: The NeoCrawler: identifying and retrieving neologisms from the internet and monitoring on-going change. *Current methods in historical semantics*, 59–98. Toim. Kathryn Allan – Justyna A. Robinson. Mouton de Gruyter, Berlin.
- Kettunen, Lauri 1949: *Hyvää vapaata suomea*. Gummerus, Jyväskylä.
- Kielikello* = *Kielikello*. Kielenhuollon tiedotuslehti. Kotimaisten kielten tutkimuskeskus, Helsinki.
Kielikellon avoin verkkoarkisto osoitteessa <http://www.kielikello.fi>.
- Kiviniemi, Eero 2006: *Suomalaisten etunimet*. SKST 1103. SKS, Helsinki.
- KKK = Kotimaisten kielten keskus
- KKK: Sanakirjoja = Sanakirjoja. Verkkosivu osoitteessa <http://www.kotus.fi/index.phtml?s=572> (viitattu 14.11.2011). KKK, Helsinki.
- KOK = *Kielitoimiston oikeinkirjoitusopas* 2007. 2., korjattu painos. Päätoim. Salli Kankaanpää. KKTJK 147. KKTJK, Helsinki.
- KKTJK = Kotimaisten kielten tutkimuskeskus
- KKTJKJ = Kotimaisten kielten tutkimuskeskuksen julkaisuja
- Klemettinen, Riina 2012: Kokemuksia lääketieteen kielen tarkistamisesta. *Kielikello* (3), 11–13.
- Koivisto, Vesa 2006: Ison suomen kieliopin termimuutoksia ja -ratkaisuja. *Kielioppi koulussa*, 25–40. Toim. Minna Harmanen – Mari Siirainen. ÄOL:n vuosikirja XLX. ÄOL, Helsinki.
- 2013: *Suomen sanojen rakenne*. SKS, Helsinki.
- Kolehmainen, Leena 2006: Saksan ja suomen verbien morfologiaa, fraseologiaa, semantiikkaa ja valenssin muutoksia. *Vir*. Verkkolektio osoitteessa http://www.kotikielenseura.fi/virittaja/hakemistot/jutut/kolehmainen2_2006.pdf (viitattu 14.12.2012).
- Kolehmainen, Taru 2011: Nominatiivi vai genetiivi yhdyssanan alussa? *Kielikello* (2), 14–17.
- 2014: *Kielenhuollon juurilla. Suomen kielen ohjailun historiaa*. KKKJ 174. SKS, Helsinki.
- Koski, Mauno 1978: Suomen sananjohdon perustyyppit. *Papers from the conference on general linguistics. Seili* 24–25.8.1978, 103–117. SKYJ 2. SKY, Turku.
- 1981: Mitä leksikaalistuminen on? *Sananmuodostuksen ongelmia. Seminaari Seilissä* 3.–4.9.1981, 5–28. SKYJ 7. Turun yliopisto, Turku.
- 1982: *Suomen johto-opin morfologiaa*. Fennistica 4. Finska institutionen vid Åbo Akademi, Åbo.
- Koskinen, Arja 1992: *Deonttisuus saamessa ja suomessa. Vertaileva tutkimus pakon, tahdon, luvan ja kyvyn modaalisuuden ilmaisemisesta saamessa ja suomessa*. Lisensiaatintyö. Suomalais-ugrilainen kielentutkimus, Suomalaisen ja yleisen kielitieteen laitos, Turun yliopisto, Turku.
- Kosonen, Kirsti 1985: *Nominatiivialkuisten yhdyssubstantiivien semantiikkaa*. Pro gradu -tutkielma. Suomen kieli, Joensuun yliopisto.
- Kotilainen, Lari 2007: *Konstruktioiden dynamiikkaa*. Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos, Helsinki. Saatavissa myös: <https://helda.helsinki.fi/bitstream/>

- handle/10138/19645/konstruk.pdf.
- Kotilainen, Lari – Varteva, Annukka 2005: Kielipoliisiainosaston yhteensijoittamisohjeet. *Kielipoliisin käsikirja*, 89–108. Toim. Pirkko Muikku-Werner. Tammi, Helsinki.
- Koukkunen, Kalevi 1990: *Atomi ja missi. Vierassanojen etymologinen sanakirja*. WSOY, Helsinki.
- Kryukova, Irina 2008: Russian advertising names as a matter of linguistic fashion. *Onoma* 43, 397–410.
- KS = *Kielitoimiston sanakirja* 2.0. Päätoim. Eija-Riitta Grönros. KTK/KKK – Kielikone Oy, Helsinki. Sähköisenä 2004, 2008 ja 2012, painettuna 2006.
- Kulttuurin rikastuminen = Otsikoimaton keskustelupuheenvuoro viestiketjussa ”Kulttuurin rikastuminen” verkko-osoitteessa <http://murobbs.muropaketti.com/threads/kulttuurin-rikastuminen.492893/page-2401> (julkaistu 26.3.2008, poimittu 10.5.2011).
- Kurki, Tommi 2005: *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen. Kielenmuutosten seuraamisesta ja niiden tarkastelussa käytettävistä menetelmistä*. SKS, Helsinki.
- Kurumada, Chigusa – Meylan, Stephan C. – Frank, Michael C. 2013: Zipfian frequency distributions facilitate word segmentation in context. *Cognition* 127, 439–453. Saatavissa myös: <http://langcog.stanford.edu/papers/KMF-cognition2013.pdf>.
- Kuusi, Osmo – Hiltunen, Elina 2011: The signification process of the future sign. *Journal of Futures Studies* 16 (1), 47–66.
- Kytömäki, Leena 1990: Nominikantaisten verbijohdosten rakennemalleja. *SJ* 34, 69–75.
- Laakkonen, Outi – Laakso, Minna 2001: Wernicken afasiaan liittyvät sanavääritymät ja niiden nonverbaali käsittely keskustelupuheessa kahden vuoden seurannan aikana. *Puhe ja kieli* 21 (4), 163–176.
- Laaksonen, Kaino 1984: Yhdyssanavirheet kuriin! *Vir.* 88 (4), 509–517.
- Laalo, Klaus 1982: Nykysuomen nominivartalotyyppien ja niiden ikäkerrostumien frekvensseistä. *Vir.* 86 (1), 22–42.
- Laine, Matti – Virtanen, Patrik 1996: Julkaisematon tekstikorpus, joka sisältää Turun Sanomien numerot ajalta 1.3.1994–31.6.1996. Kognitiivinen neurotiede, psykologian oppiaine, käyttäytymistieteiden ja filosofian laitos, Turun yliopisto.
- Laine, Päivi 2007: *Suomi tiellä sivistyskieleksi. Suomenkielisen maantieteen sanaston kehittyminen ja kehittäminen 1800-luvulla*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 77, Turun yliopisto, Turku.
- Laippala, Veronika – Ginter, Filip 2014: *Suomenkielinen Internet tekstiaineistoksi. Miljardin sanan syntaksianalyysi ja entäs sitten?* Esitelmä Kielitieteen päivillä 9.5.2014. Kielitieteen päivät 8.–10.5.2014. Turun yliopisto, Turku. Abstrakti saatavissa verkko-osoitteesta http://www.utu.fi/fi/sivustot/ktp2014/kutsupuhujat/Sivut/laippala_ginter_abs.aspx.
- Lakoff, George 1987: *Women, fire and dangerous things*. University of Chicago Press, Chicago.
- Langacker, Ronald W. 1987: *Foundations of cognitive grammar. Volume I: Theoretical prerequisites*. Stanford University Press, Stanford.
- 1988: An overview of cognitive grammar. *Topics in Cognitive linguistics*, 3–48. Toim. Brygida Rudzka-Ostyn. John Benjamins, Amsterdam.
- 1991: *Concept, image, and symbol. The cognitive basis of grammar*. Mouton de Gruyter, Berlin.
- 1999: *Grammar and conceptualization*. Mouton de Gruyter, Berlin.
- Lantee, Anna 2010: *Pitkät yhdyssubstantiivit Helsingin Sanomissa*. Pro gradu -tutkielma. Suomen kieli, Tampereen yliopisto, Tampere. Saatavissa myös: <http://tampub.uta.fi/handle/10024/81743>.
- Laudanna, A. – Burani, C. 1995: Distributional properties of derivational affixes: Implications for processing. *Morphological aspects of language processing*, 345–364. Toim. L. B. Feldman. Erlbaum, Hillsdale, NJ.
- Lauri, Ritva – Ono, Tsuyoshi 2005: Data is data and model is model: You don't discard the data that doesn't fit your model! *Language* 81 (1), 218–225.
- Lees, R. B. 1966: On a transformational analysis of compounds: A reply to Hans Marchand. *Indogermanische Forschungen* 71, 1–13.
- 1970: Problems in the grammatical analysis of English nominal compounds. *Progress in Linguistics*, 174–186. Toim. M. Bierwisch – K. E. Heidolph. Mouton, Hague.
- Lehecka, Tomas 2012: *Interrelaterande lexikala egenskaper. Engelska adjektiv importer i en svensk tidningskorpus*. Finska, finskugriska och nordiska institutionen, Humanistiska fakulteten, Helsingfors universitet, Helsingfors. Saatavissa myös: <https://helda.helsinki.fi/>

- handle/10138/37590.
- Lehrer, Adrienne 1996: Why neologisms are important to study? *Lexicology* 2 (1), 63–73.
- Lehtinen, Raija 1984: Omastatunnosta omatuntoon? *Vir.* 88 (4), 517–518.
- Lehtinen, Marja 1997: CD-Perussanakirja – uuden ajan sanasampo. *Kielikello* (1), 22–24.
- Leino, Jaakko 2003: *Antaa sen muuttua. Suomen kielen permissiivirakenne ja sen kehitys*. SKS, Helsinki.
- Leino, Pentti 1992: Ylioppilasaineiden kielivirheet: Mistä ne kertovat? *Ylioppilasaineita 1992*, 170–188. Toim. Anita Julin – Kaija Parko – Pirkko Riuraniemi – Sakari Viinamäki. SKS – ÄOL, Helsinki.
- 2002a [1987]: Kieli ja maailman hahmottaminen. *Pentti Leino. Mittoja, muotoja ja merkityksiä*, 519–548. Toim. Liisa Tainio – Aki Ontermä – Tapani Kelomäki – Minna Jaakola. SKST 868. SKS, Helsinki. Ilmestynyt alun perin vuonna 1987 teoksessa *Kieli, kertomus, kulttuuri*, 26–51. Toim. Tommi Hoikkala. Gaudeamus, Helsinki.
- 2002b [1989]: Miten tutkia suomen kieltä Suomessa? *Pentti Leino. Mittoja, muotoja ja merkityksiä*, 353–376. Toim. Liisa Tainio – Aki Ontermä – Tapani Kelomäki – Minna Jaakola. SKST 868. SKS, Helsinki. Ilmestynyt alun perin vuonna 1989 kausijulkaisussa *Specimina Fennica* I, 17–37.
- Leskinen, Heikki 1983: Otsikkokielen yhdyssanoja ja lyhenteitä. *Vir.* 87 (4), 570–572
- Levi, J. N. 1978: *The syntax and semantics of complex nominals*. Academic Press, New York.
- Levinson, S. 2000: *Presumptive meanings. The theory of generalized conversational implicature*. The MIT Press, Cambridge, Massachusetts.
- Levonen, Suvi 2006: *Verbaaliyhdyssanat HS 2000–2001 -korpuksessa*. Pro gradu -tutkielma. Suomen kieli, Tampereen yliopisto, Tampere. Saatavissa myös: <https://tampub.uta.fi/handle/10024/93744>.
- Lieber, Rochelle 1992: *Deconstructing morphology. Word formation in syntactic theory*. University of Chicago Press, Chicago.
- 2009: A lexical semantic approach to compounding. *OHC*, 78–104.
- Lieber, Rochelle – Štekauer, Pavol 2009: Introduction: status and definition of compounding. *OHC*, 3–18.
- Liljestränd, Birger 1993: *Så bildas orden. Handbok i ordbildning*. Studentlitteratur, Lund.
- Lindén, Krister – Pirinen, Tommi 2009: Weighted finite-state morphological analysis of Finnish compounding with HFST-LexC. *Proceedings of the 17th nordic conference of computational linguistics NODALIDA 2009*, 89–95. Toim. Kristiina Jokinen – Eckhard Bick. NEALT Proceedings Series 4. Northern European Association for Language Technology. Saatavissa myös: <http://dspace.utlib.ee/dspace/bitstream/handle/10062/9733/paper29.pdf>.
- Lohmann, Arne 2013: Is tree hugging the way to go? Classification trees and random forests in linguistic study. Verkkoartikkeli osoitteessa http://anglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/weitere_Uploads/Views/VIEWS_22_2013_Lohmann.pdf (viitattu 27.10.2014). *Views. Vienna English Working Papers* 22, 1–15.
- Luojola, Teemu 2006: *Kielitieteellisen aineiston kvantitatiiviset analyysimenetelmät*. Verkkoartikkeli osoitteessa http://www.ling.helsinki.fi/~fkarlss0/methods/kvant_men.pdf (muutettu 14.12.2006; viitattu 14.11.2011). Yleisen kielitieteen laitos, Helsingin yliopisto.
- Lönnrot, Elias 1958 [1866–1880]: *Suomalais-Ruotsalainen Sanakirja I–II*. Kolmas, manulmenetelmällä jäljennetty painos. WSOY, Porvoo–Helsinki.
- Malmivaara, Terhi 2004: *Luupää, puupää, puusilmä*. Näkymiä sananmuodostuksen analogisuuteen ja bahuviirihdyssanojen olemukseen. *Vir.* 108 (3), 347–363.
- Mandler, Jean M. 1984: *Stories, scripts, and scenes: Aspects of schema theory*. Erlbaum, Hillsdale, N. J.
- Mannermaa, Mika 2004: *Heikoista signaaleista vahva tulevaisuus*. WSOY, Helsinki.
- Manning, Christopher D. 2002: Probabilistic syntax. *Probabilistic linguistics*, 289–341. Toim. Rens Bod – Jennifer Hay – Stefanie Jannedy. MIT Press, Cambridge, MA.
- Marchand, H. 1967: Expansion, transposition, and derivation. *La Linguistique* 1, 13–26.
- van Marle, Jaap 1986: The Domain Hypothesis: The Study of Rival Morphological Processes. *Linguistics* 24 (3), 601–627.
- 1992: The relationship between morphological productivity and frequency: a comment on Baayen's performance-oriented conception of morphological productivity. *Yearbook of Morphology* 1991, 151–163. Toim. Geert Booij – Jaap van Marle. Kluwer, Dordrecht.
- Marslen-Wilson, William – Tyler, Lorraine – Waksler, Rachel – Older, Lynn 1994: Morphology and meaning in the English mental lexicon. *Psychological Review* 101 (1), 3–33.

- Matracki, Ivica Pesa 2006: Word formation tendencies in contemporary Italian. *Studia Romanica et Anglica Zagrabiensia* 51, 103–146.
- Matthews, Peter Hugoe 1991 [1974]: *Morphology. Second edition*. Second, updated and revised edition. Cambridge University Press, Cambridge.
- McCarthy, Philip – Jarvis, Scott 2010: MTL, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods* 42 (2), 381–392. Saatavissa myös: <http://dx.doi.org/10.3758/BRM.42.2.381>.
- McEnery, Tony – Hardie, Andrew 2012: *Corpus linguistics: Method, theory and practice*. Cambridge University Press, Cambridge.
- McEnery, Tony – Wilson, Andrew 1996: *Corpus linguistics*. Edinburgh University Press, Edinburgh.
- Meibauer, Jörg 2007: How marginal are phrasal compounds: Generalized insertion, expressivity, and I/Q-interaction. *Morphology* 17 (2), 233–259.
- Mendoza-Denton, Norma – Hay, Jennifer – Jannedy, Stephanie 2003: Probabilistic sociolinguistics: Beyond variable rules. *Probabilistic linguistics*, 97–138. Toim. Rens Bod – Jennifer Hay – Stephanie Jannedy. Massachusetts Institute of Technology Press, Cambridge MA.
- Metropoli: A. H. pyysi S/M-seksiä = Metropoli-sivuston Facebook-päivitys verkko-osoitteessa https://www.facebook.com/permalink.php?story_fbid=10150514993083135&id=227828768134 (julkaistu 29.11.2011, poimittu 6.6.2014).
- Mielikäinen, Aila 1977: Yhdyssanat ja sanaliitot sekaisin. *Vir.* 81 (4), 468.
- Mikkola, Eino 1971: *Das Kompositum: Eine vergleichende Studie über die Wortzusammensetzung im Finnischen und in den indogermanischen Sprachen I: Der Anteil der Komposita an der Sprache des modernen Romans*. 2. painos. Sprache und Denken: Finnische Beiträge zur Philosophie und Sprachwissenschaft V, Helsinki.
- Mikkonen, Pirjo 2002: Uusitalon vai Uudentalon? *Kielikello* (1), 38.
- Mikone, Eve 2002: *Deskriptiiviset sanat. Määritelmät, muoto ja merkitys*. SKS, Helsinki.
- Moscoso del Prado Martín ym. 2004 = Moscoso del Prado Martín, Fermín – Bertram, Raymond – Häikiö, Tuomo – Schreuder, Robert – Baayen, R. Harald 2004: Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30 (6), 1271–1278.
- Mueller, J. H. – Schuessler, K. F. 1961: *Statistical reasoning in sociology*. Houghton Mifflin, Boston.
- Multanen, Eeva 1991: *Metsäprofeetasta volyymiajatteluun. Tilapäisten yhdyssubstantiivien semantiikkaa*. Laudaturtyö. Tampereen yliopiston suomen kielen laitos, Tampere.
- Mustajoki, Arto 2009: Miksi tieteen kieli on epätarkkaa? *Vuosikirja 2009 Academia Scientiarum Fennica, Suomalaisen tiedeakatemian vuosikirja*, 65–68. Suomalainen Tiedeakatemia, Helsinki. Saatavissa myös: <http://hdl.handle.net/10138/24828>.
- 2012: A speaker-oriented multidimensional approach to risk and causes of miscommunication. *Language and Dialogue* 2 (2), 216–243.
- Mäkisalo, Jukka 2000a: *Grammar and experimental evidence in Finnish compounds*. Studies in Language 35. Faculty of Humanities, University of Joensuu, Joensuu.
- 2000b: Miten yhdyssanat ovat leksikon ja kieliopin rajapinnalla? Väitöksenalkajaisitelmä Joensuun yliopistossa 21. kesäkuuta 2000. *Vir.* 104 (4), 576–579.
- Napsu: Mitä eroa on kosmonautilla ja astronautilla? = Otsikoimaton vastaus keskustelupuheenvuoroon ”Mitä eroa on kosmonautilla ja astronautilla” verkko-osoitteessa <http://www.napsu.fi/vastaukset/kysymys/30761> (julkaistu 2.4.2010, poimittu 10.5.2011).
- Nenonen, Marja 2002: *Idiomit ja leksikko. Lausekeidiomien syntaktisia, semanttisia ja morfologisia piirteitä suomen kielessä*. Joensuun yliopiston humanistisia julkaisuja 29. Joensuun yliopisto, Joensuu.
- NES = Häkkinen, Kaisa 2004: *Nykysuomen etymologinen sanakirja*. WSOY, Helsinki.
- Newmeyer, Frederick J. 2003: Grammar is grammar and usage is usage. *Language* 79 (4), 682–707.
- 2005: A reply to the critiques of ‘Grammar is grammar and usage is usage’. *Language* 81 (1), 229–236.
- Niemi, Jarkko 2010: Myönnytelymuotti – erimielisyyttä enteilevä samanmielisyyden konstruktio. *Vir.* 114 (2), 196–222.
- Niemi, Jussi 1999: Production of grammatical number in specific language impairment: An elicitation experiment on Finnish. *Brain and Language* 68 (1–2), 262–267.

- Niemi, Jussi 2009: Compounds in Finnish. *Lingue e Linguaggio* 8 (2), 237–256.
- Niemi, Jussi – Heikkinen, Janne – Järvikivi, Juhani 2001: Miksei kania vaikka munoja? *XXVII Kielitieteen päivät Oulussa* 19.–20.5.2000, 191–197. Toim. Helena Sulkala – Leena Nissilä. Acta Universitatis Ouluensis, B Humaniora, 41. Oulun yliopisto, Oulu.
- Niemikorpi, Antero 1990: *Suomen kielen sanaston frekvenssianalyysia. Frequency analysis of Finnish vocabulary*. Vaasan korkeakoulun julkaisuja. Tutkimuksia No 150. Kielitiede 18. Vaasan korkeakoulu, Vaasa.
- 1991: *Suomen kielen sanaston dynamiikkaa*. Acta Wasaensia 26. Kielitiede 2. Vaasan yliopisto, Vaasa.
- 1996: *Liekepostista tuikeilmaisimeen ja sulhasesta kuraankeliin. Erikoiskielen rakenteellisesta ja tyylillisestä vaihtelusta*. Vaasan yliopiston julkaisuja. Tutkimuksia 213. Kielitiede 34. Vaasan yliopisto, Vaasa.
- Nieminen, Tommi 1996: *Suomen kielen puherytmi*. Pro gradu -tutkielma. Suomen kielen ja yleisen kielitieteen laitos, Tampereen yliopisto.
- Nieminen, Tommi – Lantee, Anna 2011: Yhdyssanasta suomen kielessä ja puheessa. *XXVI Fonetiikan päivät 2010*, 69–74. Toim. Stefan Werner – Tomi Kinnunen. Verkkojulkaisu osoitteessa <http://urn.fi/URN:ISBN:978-952-61-0391-4> (viitattu 21.9.2012). University of Eastern Finland electronic publications, University of Eastern Finland, Joensuu.
- Niinimäki, Anneli 1991: *Yhdyssanojen hahmotus ja oikeinkirjoitus peruskoulun päättävien oppilaiden teksteissä*. Lisensiaatintyö. Suomen kielen laitos, Tampereen yliopisto, Tampere.
- 1992: Sanaliittojen tiivistyminen yhdyssanaksi. *Vir.* 96 (2–3), 283–286.
- Nikanne, Osmo 1971: ”Sanan käyttötapa”. Lisää yhdyssanan attribuutista. *Vir.* 75 (1), 112–114.
- Nikkilä, Osmo 1994: *Loppuheitto ja vanha kirjasuomi. Suomen kielen i:n loppuheiton historiaa*. Opera Fennistica & Linguistica 8. Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos, Tampere.
- Nikolaev, Alexandre 2002: Eräiden suomen taivutustyyppien produktiivisuudesta. *Puhe ja kieli* 22 (3), 113–124. Saatavissa myös: <http://ojs.tsv.fi/index.php/pk/article/view/8065/6164>.
- 2011: *Suomen nominien taivutusjärjestelmän produktiivisuudesta: Korpuslingvistisiä ja kokeellisia tutkimuksia*. Filosofinen tiedekunta, Itä-Suomen yliopisto, Joensuu.
- Nikolaev, Alexandre – Niemi, Jussi 2005: Suomen nominien taivutuksesta: rytmi-, sivupainoja agglutinaatiohypoteesien testausta. *Vir.* 209 (4), 482–505.
- 2006: Nominien paradigmaattistuminen suomessa. Millä rakenteellisilla ehdoilla kielenkäyttäjät sitovat potentiaaliset nominit taivutusluokkiin? *Vir.* 110 (1), 46–69.
- 2008: Suomen nominien taivutusjärjestelmän produktiivisuuden indekseistä. *Vir.* 112 (4), 518–544.
- NS = *Nyky-suomen sanakirja I–VI*. Valtion toimeksiannosta teettänyt SKS. WSOY, Porvoo–Helsinki 1951–1961. Lista hakusanoista sanaluokka- ja taivutustietoineen saatavissa myös sähköisenä. KKK, Helsinki. Käyttö vaatii luvan.
- NSST = KKK:n Nyky-suomen sanastotietokanta. Online-tietokanta. Päivitetään jatkuvasti. Tiedot ajalta 1.1.2000–1.9.2009, haettu 2.10.2009. KKK, Helsinki. (Sisältyy kokoelmaan Nykykielen aineistot: Nyky-suomen sanakokoelmat). Saatavuus: Arkistotietokantaa voi käyttää KKK:n intranetissä, mutta käyttö vaatii luvan.
- Nuutinen, Liisa 2012: Roheaa räimettä. Kielitoimiston sanakirjan uudissanoja. *Kielikello* (1), 7–9.
- Nuutinen, Olli 1989: Käännöslainat. *Nyky-suomen sanavarat*, 110–126. Toim. Jouko Vesikansa. WSOY, Porvoo–Helsinki–Juva.
- OED = *Oxford English dictionary* (1884–). Vuoteen 1933 asti nimellä *New English dictionary*. Oxford University Press, Oxford. Saatavissa: www.oed.com, vaatii käyttäjätunnuksen.
- OED: About = About. Verkkosivu osoitteessa <http://public.oed.com/about/> (viitattu 10.9.2014). Oxford University Press, Oxford.
- OED: Glossary = Glossary. Verkkosivu osoitteessa <http://public.oed.com/how-to-use-the-oed/glossary> (viitattu 10.9.2014). Oxford University Press, Oxford.
- OHC = *The Oxford handbook of compounding*. Toim. Rochelle Lieber – Pavol Štekauer. Oxford University Press, Oxford 2009.
- Ojansuu, Heikki 1908: Muutamia omituisia yhdyssanoja. *Vir.* 12, 103–104.
- Olsen, Susan 2001: Copulative compounds. A closer look at the interface between morphology and

- syntax. *Yearbook of Morphology* 2000, 279–320. Toim. Geert Booij – Jaap van Marle. Kluwer, Dordrecht.
- Omorfi 2012 = Omorfi – Open morphology for Finnish. Vapaan lähdekoodin tietokoneohjelmasta vuonna 2012 saatavissa ollut versio. Ohjelman verkkosivut osoitteessa <https://code.google.com/p/omorfi/>. GNU General public license.
- Omorfi 2014 = Omorfi – Free open source morphological analyser for Finnish language. Verkkosivu osoitteessa <https://code.google.com/p/omorfi/> (viitattu 20.6.2014).
- Ong, Walter J. 2002 [1982]: *Orality and literacy. The technologizing of the world*. Routledge, London – New York.
- Onikki, Tiina 1994: Skeema merkitysrakenteen ja taustatiedon kuvauksessa. *Näkökulmia polysemiaan*, 70–138. Toim. Pentti Leino – Tiina Onikki. Kieli 8. Helsingin yliopiston suomen kielen laitos, Helsinki.
- Onikki-Rantajääskö, Tiina 2001: *Sarjoja. Nykysuomen paikallissijaiset olotilanilmaukset kielen analogisuuden ilmentäjinä*. SKS, Helsinki.
- Orpana, Terttu 1988: *Kuvaus vai kommentti. Tutkimus suomen kielen adjektiivadverbien semanttisesta tulkinnasta*. Opera Fennistica & Linguistica 2. Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos, Tampere.
- Palander, Marjatta 2002: Yhdysadjektiivit – muuttuvia muoti-ilmauksia. *Kielikello* (1), 15–17.
- Pajunen, Anneli – Palomäki, Ulla 1984: *Tilastotietoja suomen kielen rakenteesta 1. Frequence analysis of spoken and written discourse in Finnish* 1. KKTKJ 30. KKTK, Helsinki.
- Party = R:n (ks. R 2014) Party-paketin (versio 1.0-15) help-tiedosto. Saatavissa myös: <http://cran.r-project.org/web/packages/party/party.pdf>.
- Paryzek, Piotr 2008: Comparison of selected methods for the retrieval of neologisms. *Investigationes Linguisticae* XVI, 163–181. Saatavissa myös: http://www.inveling.amu.edu.pl/pdf/Piotr_Paryzek_inve16.pdf.
- Paul, Herman 1898: *Prinzipien der Sprachgeschichte*. 3. Aufl. Max Niemeyer, Halle.
- Pearson, Karl 1900: On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series* 5, 50 (302), 157–175. Julkaistu sähköisenä 2009. Taylor & Francis, <http://www.tandfonline.com>.
- Pentikäinen, Ritva 1998: *Henkilöstöhallinnon termistö. Teoreettis-rakenteellinen tarkastelu*. Pro gradu -tutkielma. Suomen kielen laitos, Vaasan yliopisto, Vaasa.
- Penttilä, Aarni 1963 [1957]: *Suomen kielioppi*. 2., tarkistettu painos. Werner Söderström osakeyhtiö, Porvoo–Helsinki.
- Petrova, Oksana 2011: *Of pearls and pigs. A conceptual-semantic tiernet approach to formal representation of structure and variation of phraseological units*. Åbo Akademi University Press, Åbo. Saatavissa myös: http://www.doria.fi/bitstream/handle/10024/69852/petrova_oksana.pdf.
- Pilke, Nina 1999: Mitä täsmällistä on täsmäkuivaimessa? *Erikoiskielet ja käännösteoria. VAKKI-symposiumi XIX, Vaasa 13.-14.2.1999*, 277–288. Toim. Rune Ingo. Käännösteorian ja ammattikielten tutkijaryhmä, Vaasan yliopisto, Vaasa.
- Pirinen, Tommi A. 2011: Modularisation of Finnish finite-state language description – Towards wide collaboration in open source development of morphological analyser. *Proceedings of the 18th Nordic Conference of Computational Linguistics NODALIDA 2011*, 299–301. Toim. Bolette Sandford Pedersen – Gunta Nešpore – Inguna Skadina. NEALT Proceedings Series 11. Northern European Association for Language Technology. Saatavissa myös: http://dSPACE.utlib.ee/dSPACE/bitstream/handle/10062/17358/0/Pirinen_58.pdf.
- Pitkänen, Kaarina 2008: *Suomi kasvitieteen kieleksi. Elias Lönnrot termistön kehittäjänä*. SKST 1164. SKS, Helsinki.
- Plag, Ingo 1999: *Morphological productivity: structural constraints in English derivation*. Mouton de Gruyter, Berlin – New York.
- 2003: *Word-formation in English*. Cambridge Textbooks in Linguistics. Cambridge University Press, Cambridge.
- 2006: Productivity. *Handbook of English linguistics*, 537–556. Toim. B. Aarts – A. McMahon. Blackwell, Oxford.

- Pollatsek, Alexander – Bertram, Raymond – Hyönä, Jukka 2011: Processing novel and lexicalised Finnish compound words. *Journal of Cognitive Psychology* 23 (7), 795–810.
- Pollatsek, Alexander – Hyönä, Jukka – Bertram, Raymond 2000: The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance* 26 (2), 820–833. Saatavissa myös: http://users.utu.fi/hyona/Pollatsek%20et%20al_2000.pdf.
- Pollatsek, Alexander – Hyönä, Jukka 2005: The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes* 20 (1), 261–290. Saatavissa myös: http://users.utu.fi/hyona/pollatsek_hy%C3%B6n%C3%A4_2005.pdf.
- Popova ym. 2005 = Popova, T. V. – Ratsiburskaja L. V. – Gugunava D. V. 2005: Neologija i neografija sovremennogo russkogo jazyka. Flinta, Moskva = Попова, Т. В. – Рацибурская, Л. В. – Гугунава Д. В. 2005: *Неология и неография современного русского языка*. Флинта, Москва.
- Prčić, Tvrtko 2005: Prefixes vs initial combining forms in English: A lexicographic perspective. *International Journal of Lexicography* 18 (3), 313–334. Oxford University Press, Oxford.
- Priiki, Katri 2014: Kaakkois-Satakunnan henkilöviitteiset *hän, se, tätä ja toi* subjekteina. *SJ* 56, 86–107. *PS = Suomen kielen perussanakirja*. Valtion painatuskeskus – KKTK, Helsinki. Painettuna 1990–1994, sähköisenä 1996 (ks. CD-PS).
- Pulkkinen, Paavo 1972: *Nykysuomen kehitys*. Tietolipas 72. SKS, Helsinki.
- 1980: Väkinäinen yhdyssana vai sujuva sanaliitto? *Vir.* 84 (2), 179–181.
- 1984: *Lokarista sponsoriin. Englantilaisia lainoja suomen kielessä*. Otava, Helsinki.
- Punttila, Matti 1995: Hyperhyvää vai huippuhuonoa? Tarpeellisista ja tarpeettomista vahvistussanoista. *Sana kiertää. Kirjoituksia suomen kielestä*, 126–130. Toim. Maija Länsimäki. KTKKJ 80. KKTK, Helsinki. Ensimmäisen kerran julkaistu Kielikellossa 1990.
- Pustyl'nikov, Olga – Schneider-Wiejowski, Karina 2009: Measuring morphological productivity. *Studies in Quantitative Linguistics* 1–9 (5), 106–125. Toim. R. Köhler. RAM, Lüdenscheid. Saatavissa myös: <http://www.homes.uni-bielefeld.de/opustyl'nikov/pustyl'nikov/pdfs/Pustyl'nikovSchneider-WiejowskiV4-FINAL.pdf>.
- Pyllkkänen, L. – Feintuch, S. – Hopkins, E. – Marantz, A. 2004: Neural correlates of the effects of morphological family frequency and family size: an MEG study. *Cognition* 91, B35–B45. Saatavissa myös: https://files.nyu.edu/mp108/public/Morphofam_cognition_pub.pdf.
- R 2014 = R Development Core Team 2014: *R – A language and environment for statistical computing*. R version 3.1.1 (2014-07-10) "Sock it to Me". The R Foundation for Statistical Computing, Vienna. Saatavissa: <http://www.r-project.org/>.
- Rahtu, Toini 1984: Suomen nominaalkuiset yhdysverbit. *Vir.* 88 (4), 409–430.
- Rajakallio, Riitta 1995: *Suomalaisen atk-sanaston kehitys 1960-luvulta 1990-luvulle*. Pro gradu -tutkielma. Suomalaisen ja yleisen kielitieteen laitos, Turun yliopisto, Turku.
- Ralli, Angela 2009: Modern Greek V V dvandva compounds: A linguistic innovation in the history of the Indo-European languages. *Word Structure* 2 (1), 48–68.
- Ravila, Paavo 1961: *Johdatu kielihistoriaan*. 2., uusittu painos. Tietolipas 3. SKS, Helsinki.
- Rayner, Keith 1998: Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124 (3), 372–422.
- Rayner, Keith – Pollatsek, Alexander 1989: *Psychology of reading*. Erlbaum, Englewood Cliffs, NJ.
- Rey, Alain 1976: Néologisme: un pseudo-concept? *Cahiers de lexicologie* 28 (1), 3–17.
- 1995: *Essays on terminology*. Käänt. Juan C. Sager. J. Benjamins, Amsterdam–Philadelphia.
- Richards, Brian 2009 [1987]: Type/token ratios: what do they really tell us? *Journal of Child Language* 14 (2), 1987, 201–209. Julkaistu sähköisenä 2009, saatavissa: <http://hfroehlich.files.wordpress.com/2013/05/s0305000900012885a.pdf>.
- Rintala, Päivi 1972: *Suomen läntä-loppuiset adjektiivit*. SKST 306. SKS, Helsinki.
- 1978: Sananjohdon produktiivisuudesta ja sen rajoituksista nykysuomessa. *Juhlakirja Osmo Ikolan 60-vuotispäiväksi 6.2.1978*, 151–164. Toim. Alho Alhoniemi ym. Turun yliopiston suomalaisen ja yleisen kielitieteen julkaisuja 6. Turun Yliopistosäätiö, Turku.
- 1992: Suomen kirjakielen normeista. *SJ* 34, 47–68. Saatavissa myös: <https://helda.helsinki.fi/bitstream/handle/10224/3576/rintala47-68.pdf>.
- Rissanen, Terhi 1998: *The categories of nominals and verbals and their morphology in Finnish sign language*. Licentiate thesis in general linguistics. The department of Finnish and general

- linguistics, University of Turku.
- Rohrer, Jonathan D. – Rossor, Martin N. – Warren, Jason D. 2009: Neologistic jargon aphasia and agraphia in primary progressive aphasia. *Journal of the Neurological Science* 277 (1–2), 155–159. Saatavissa myös: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2633035/>.
- Romalo, Valeria Gutu 1992: Linguistic Fashion. *Limba romana* 41 (3), 161–166.
- Rondeau, Guy 1981: *Introduction à la terminologie*. Centre Educatif et Culturel, Québec.
- Rosch, Eleanor 1977: Classification of real-world objects: Origins and representations in cognition. *Thinking. Readings in cognitive science*, 212–222. Toim. P. N. Johnson-Laird – P. C. Wason. Cambridge University Press, Cambridge.
- RStudio 2009–2013 = RStudio, Inc. 2009–2013: *RStudio – Integrated development environment for R*. Version 0.98.490. Boston, MA. Saatavissa: <http://www.rstudio.org/>.
- Rothman, Kenneth J. 1990: No adjustments are needed for multiple comparisons. *Epidemiology* 1 (1), 43–46.
- Rumelhart, David E. – Ortony, Andrew 1977: The representation of knowledge in memory. *Schooling and the acquisition of knowledge*, 99–136. Toim. Richard C. Anderson – Rand J. Spiro – William E. Montague. Lawrence Erlbaum Associates Publishers, Hillsdale, N.J.
- Ryder, Mary Ellen 1994: *Ordered chaos. The interpretation of English noun-noun compounds*. University of California Publications in Linguistic 123. University of California Press, Berkeley.
- Rytkönen, Ahti 1937: Koloratiivinen konstruktio. *Vir.* 41 (1), 95–104.
- Räsänen, Alpo 1986: Sananmuodostus ja konteksti. *Vir.* 90, 155–163. Saatavissa myös: http://www.kotikielenseura.fi/virittaja/hakemistot/jutut/1986_155.pdf
- 1990: Yhdyssanojen merkitysoppia. *Kielikello* (4), 3–5.
- Räsänen, Matti 2010: *Päivä päivältä enemmän ja enemmän. Suomen toistokonstruktioita*. Yliopistopaino, Helsinki. Saatavissa myös: <https://helda.helsinki.fi/bitstream/handle/10138/19663/paivapai.pdf>.
- Saarimaa, E. A. 1971 [1947]: *Kielenopas*. 8. painos. Tarkistanut Paavo Pulkkinen. WSOY, Porvoo–Helsinki.
- Sadeniemi, Matti 1942: Nominatiivi - genetiivi? *Vir.* 46 (2), 197–201.
- 1949: *Metriikkamme perusteet ja sovellutusta moderneihin ja antiikin mittoihin*. SKST 236. SKS, Helsinki.
- 1955: ”Pölysidottu tie”. *Vir.* 59 (3), 299–300.
- Saukkonen, Pauli 1973: Suomen kielen yhdyssanojen rakenne. *Commentationes Fenno-Ugricae in honorem Erkki Itkonen sexagenerii die XXVI mensis aprilis anno MCMLXXXIII*, 332–339. Suomalais-ugrilaisen seuran toimituksia 150. Suomalais-ugrilainen seura, Helsinki.
- Scalise, Sergio 1992: Compounding in Italian. *Rivista di Linguistica* 4, 175–199.
- Scalise, Sergio – Guevara, Emiliano 2005: The lexicalist approach to word-formation and the notion of the lexicon. *Handbook of word-formation*, 147–188. Toim. Pavol Štekauer – Rochelle Lieber. Studies in Natural Language and Linguistic Theory 64. Springer Netherlands, Dordrecht.
- Schmidt, Günter Dietrich 1987: Das Affixoid. Zur Notwendigkeit und Brauchbarkeit eines beliebten Zwischenbegriffs in der Wortbildung. *Deutsche Lehnwortbildung. Beiträge zur Erforschung der Wortbildung mit entlehnten WB-Einheiten im Deutschen*, 53–101. Toim. Gabriele Hoppe – Alan Kirkness – Elisabeth Link – Isolde Nortmeyer. Niemeyer, Tübingen.
- Schreuder, R. – Baayen, R. H. 1997: How complex simplex words can be. *Journal of Memory and Language* 37, 118–139.
- Seppänen, Lauri 1981: Nominaalikompositumien semantiikkaa. *Sananmuodostuksen ongelmia. Seminaari Seilissä* 3.–4.9.1981, 69–82. SKYJ 7. Turun yliopisto, Turku.
- Setälä, E. N. 1898: *Suomen kieloppi. Äänne- ja sanaoppi oppikoulua ja omin päin opiskelua varten*. Otava, Helsinki.
- Shapiro, Samuel Sandford – Wilk, Martin 1965: An analysis of variance test for normality (complete samples). *Biometrika* 52 (3–4), 591–611.
- Sinisalo, Aino 1987: Uudellamaalla, Vanhassakaupungissa ja Pitkälläsillalla. *Kielikello* (4), 8–13. *SJ = Sananjalka*. Suomen Kielen Seuran vuosikirja. Suomen Kielen Seura, Turku.
- Sjöblom, Paula 2006: *Toiminimen toimenkuva. Suomalaisen yritysnimistön rakenne ja funktiot*. SKS, Helsinki.
- SKS = Suomalaisen Kirjallisuuden Seura

- SKST = Suomalaisen Kirjallisuuden Seuran toimituksia
 SKY = Suomen kielitieteellinen yhdistys
 SKYJ = Suomen kielitieteellisen yhdistyksen julkaisuja
 SMS = *Suomen murteiden sanakirja* 1–8. KKTKJ 36. KKTK–Valtion painatuskeskus/Edita–SKS, Helsinki 1985–2012.
- Sounio, Lisa 2010: Lapsiperheen arki täynnä tunnetta & taidetta. Blogikirjoitus verkko-osoitteessa <http://liskasounio.typepad.com/blogi/2010/04/lasten-ehdoilla.html> (julkaistu 18.4.2010, poimittu 20.12.2011).
- Spencer, Andrew 1991: *Morphological theory. An introduction to word structure in generative grammar*. Blackwell, Oxford.
- 2003: Morphology. *The handbook of linguistics*, 213–237. Toim. Mark Aronoff – Janie Rees-Miller. Wiley–Blackwell, Oxford.
- Stats = R:n (ks. R 2014) Stats-peruspaketin (versio 3.1.1) help-tiedosto.
- Štekauer, Pavol 2001: Beheading the word? Please, stop the execution. *Folia Linguistica* 34 (3–4), 333–355.
- 2002: On the theory of neologisms and nonce-formations. *Australian Journal of Linguistics* 22 (1), 97–112.
- 2009: Meaning predictability of novel context-free compounds. *OHC*, 272–297.
- Stichauer, Pavel 2009: On the productivity of (not only) verb-noun compounds in Italian. *Casopis pro Moderni Filologii* 91 (2), 65–77.
- Strobl, Carolin – Hothorn, Torsten – Zeileis, Achim 2009: ”Party on!” Verkkoartikkeli osoitteessa http://journal.r-project.org/archive/2009-2/RJournal_2009-2_Strobl-et-al.pdf (viitattu 24.10.2014). *The R Journal* 1 (2), 14–17.
- Strobl, Carolin – Malley, James – Tutz, Gerhard 2009: An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14 (4), 323–348.
- Stubbs, Michael 2001: *Words and phrases. Corpus studies of lexical semantics*. Blackwell, Oxford.
- Student 1908: The probable error of a mean. *Biometrika* 6 (1), 1–25. Saatavissa myös: http://www.aliquote.org/cours/2012_biomed/biblio/Student1908.pdf.
- Suomi24: Kosmonautti = Otsikoimaton vastaus keskustelupuheenvuoroon ”Kosmonautti, astronautti, taikonautti...” verkko-osoitteessa <http://keskustelu.suomi24.fi/node/631574> (julkaistu 16.10.2003, poimittu 10.5.2011).
- Suomi24: Mikä ekaks = Keskustelupuheenvuoro ”Mikä ekaks auto maatti aseeksi alle 150e” verkko-osoitteessa <http://keskustelu.suomi24.fi/node/5644646> (julkaistu 21.4.2008, poimittu 9.9.2014).
- Suomi, Kari 1985: On detecting words and word boundaries in Finnish: A survey of potential word boundary signals. *Nordic Journal of Linguistics* 8 (2), 211–231.
- Säily, Tanja 2008: Productivity of the suffixes *-ness* and *-ity* in 17th-century English letters: a sociolinguistic approach. Pro Gradu Thesis. Department of English, University of Helsinki. Saatavissa myös: <https://helda.helsinki.fi/bitstream/handle/10138/19323/producti.pdf>.
- 2014: *Sociolinguistic variation in English derivational productivity. Studies and methods in diachronic corpus linguistics*. Société Néophilologique, Helsinki. Saatavissa myös: <https://helda.helsinki.fi/bitstream/handle/10138/136128/sociolin.pdf>.
- Taylor, John R. 1998: Syntactic constructions as prototype categories. *The new psychology of language: Cognitive and functional approaches to language structure*, 177–202. Toim. M. Tomasello. Lawrence Erlbaum, Mahwah, NJ.
- Templin, M. C. 1957: *Certain language skills in children*. University of Minnesota Press, Minneapolis.
- TTP: Kielitiede = Tieteen termipankki. Kielitiede. Verkkosivu osoitteessa <http://tieteentermipankki.fi/wiki/Kielitiede> (viitattu 26.03.2014). Tieteen termipankki – Creative Commons.
- TSK: Millä perusteella termisuosituksia = Millä perusteella termisuosituksia annetaan? Verkkosivu osoitteessa http://www.tsk.fi/tsk/termitalkoot/mill%C3%A4_perusteella_termisuosituksia_annetaan-443.html (viitattu 16.1.2014). Sanastokeskus TSK, Helsinki.
- TT = Tyysteri, Laura 2009–2015: Tutkimustietokanta. NSST:sta 1.1.2000–1.9.2009 poimittujen tietojen pohjalta koostettu julkaisematon yhdyssanatiekanta. Suomen kieli, Turun yliopisto, Turku.
- Tuomi, Tuomo 1989: Yleiskielemme murrepohjainen sanasto. *Nyky-suomen sanavarat*, 28–56. Toim.

- Jouko Vesikansa. WSOY, Helsinki.
- Tuomi, Tuomo 1995: Täpötäysi ja patajuoppo. *Sana kiertää. Kirjoituksia suomen kielestä*, 154–157. Toim. Maija Länsimäki. KTKKJ 80. KTKK, Helsinki. Ensimmäisen kerran julkaistu Kielikellossa 1991.
- Tuomikoski, R. 1973: Adjektiivin ja adverbien erosta. *Vir.* 77 (3), 199–214.
- Tyysteri, Laura 2009: *Erikoiskielen yhdyssanojen rakenne erikoiskielen erityisluonteen kuvaajana*. Pro gradu -tutkielma. Suomalaisen ja yleisen kielitieteen laitos, Turun yliopisto, Turku. Saatavissa myös: <http://www.doria.fi/bitstream/handle/10024/46941/gradu2009tyysteri.pdf>.
- 2011a: Miksei sinkkuvaimo vakiinnu? *Turun Sanomat* 30.11.2011, 22.
- 2011b: Möhköfantin anatomia. Esitelmä Kielitieteen päivillä Joensuun yliopistossa 26.5.2011.
- 2011c: Sukupuolen merkitseminen *Nykysuomen sanakirjan ja Kielitoimiston sanakirjan nais- ja mies-alkuisissa yhdyssubstantiiveissa*. *SJ* 53, 162–180.
- UD = *Urban Dictionary*. Urban Dictionary LLC 1999–2014 (päivitetään edelleen). Verkkosanakirja osoitteessa <http://www.urbandictionary.com/>.
- Ulfsbjorninn, Shanti 2009: The internal structure of compounds: A phase account of aphasia. *Iberia. International Journal of Theoretical Linguistics* 1 (1), 143–168. Saatavissa myös <http://www.siff.us.es/iberia/index.php/ij/article/viewFile/7/9>.
- US = *Urbaani sanakirja*. Verkkosanakirja osoitteessa <http://urbaanisankirja.com/>. Nixarn, Turku 2007–2014 (päivitetään edelleen).
- Vainio, Seppo – Hyönä, Jukka – Pajunen, Anneli 2003: Facilitatory and inhibitory effects of grammatical agreement: evidence from readers' eye fixation patterns. *Brain and Language* 85 (2), 197–202. Saatavissa myös: http://users.utu.fi/hyona/Vainio%20et%20al_B&L_2003.pdf.
- 2008: Processing modifier-head agreement in reading: Evidence for a delayed effect of agreement. *Memory & Cognition* 36 (2), 329–340.
- Vaittinen, Tanja 2003: Vanhan kirjasuomen yhdysverbit. *SJ* 45, 45–66.
- 2007: Mikael Agricolan teosten yhdysadjektiivit. *Vir.* 111 (4), 523–542.
- Wang, Chamont 1993: *Sense and nonsense of statistical inference. Controversy, misuse, and subtlety*. Marcel Dekker, New York.
- Vanhatalo, Pirkko 1998: *Sanaliitto vai yhdyssana*. Lisensiaatintyö. Suomen kielen laitos, Vaasan yliopisto, Vaasa.
- Welch, B. L. 1947: The generalization of "Student's" problem when several different population variances are involved. *Biometrika* 34 (1), 28–35. Saatavissa myös: <http://www.dcs.cience.net/welch-1947.pdf>.
- 1951: On the comparison of several mean values: an alternative approach. *Biometrika* 38 (3), 330–336. Saatavissa myös: <http://www.soph.uab.edu/Statgenetics/People/MBeasley/Courses/Welch1951.pdf>.
- Vesikansa, Jouko 1989: Yhdyssanat. *Nykysuomen sanavarat*, 213–258. Toim. Jouko Vesikansa. WSOY, Helsinki.
- Viinikainen, Taru 2010: *Taipuuko "akrobaatti Aleksandra"? Nimikekonstruktio ja nimikkeen taipuminen lehtikielessä 1900-luvulta 2000-luvulle*. Publications of the University of Eastern Finland. Dissertations in Education, Humanities, and Theology 1. Itä-Suomen yliopisto, Joensuu. Saatavissa myös: http://epublications.uef.fi/pub/urn_isbn_978-952-61-0003-6/urn_isbn_978-952-61-0003-6.pdf.
- Wikipedia FI = *Wikipedia, vapaa tietosanakirja*. Verkkotietosanakirja osoitteessa <http://fi.wikipedia.org> (viitattu 20.9.2012). Wikimedia Foundation – Creative Commons.
- Williams, E. 1981: On the notions "lexically related" and "head of a word". *Linguistic Inquiry* 12 (2), 245–274.
- Wilming ym. 2011 = Wilming, Niklas – Betz, Torsten – Kietzmann, Tim C. – König, Peter 2011: Measures and limits of models of fixation selection. *PLoS ONE* 6 (9), 1–19.
- Vir.* = *Virittäjä*. Kotikielen Seura, Helsinki.
- VISK = *Ison suomen kielipiipin verkkoversio*. Hakulinen, Auli – Vilkuna, Maria – Korhonen, Riitta – Koivisto, Vesa – Heinonen, Tarja Riitta – Alho, Irja. Verkkosivusto osoitteessa <http://scripta.kotus.fi/visk>. KTKK, Helsinki 2008.
- Wittgensteinin, Ludwig 1981 [1953]: *Filosofisia tutkimuksia*. Suom. Heikki Nyman. WSOY, Porvoo–Helsinki–Juva.

- Voikko = Voikko, Free linguistic software and data for Finnish. Lisätietoja ja linkkejä lataussivustoille verkko-osoitteessa <http://voikko.puimula.org/> (viitattu 3.12.2014).
- Voikko: UKK = Usein (tai harvemmin) kysyttyjä kysymyksiä. Verkkosivu osoitteessa <http://voikko.puimula.org/faq.html> (viitattu 3.12.2014).
- Ylikulju, Marja 2004: Helppoja ja vaikeita yhdyssanoja. *Kielikello* (1), 13–17.
- Yoon, Jiyoung 2011: Productivity of Spanish verb-noun compounds: Patterns of metonymy and metaphor. *Review of Cognitive Linguistics* 9 (1), 83–106.
- Zipf, George Kingsley 1965: *The psycho-biology of language: an introduction to dynamic philology*. MIT Press, Cambridge, Mass.
- Zwicky, A. 1985: Heads. *Journal of Linguistics* 21, 1–29. Saatavissa myös: <http://web.stanford.edu/~zwicky/heads.pdf>.
- ÄOL = Äidinkielen opettajien liitto
- Östman, Jan-Ola 1988: Konstruktiokielioppi. *Suomen kielitieteellisen yhdistyksen vuosikirja* 1988, 173–187. Toim. Jan-Ola Östman – Urpo Nikanne – Kristiina Jokinen – Kalevi Wiik. SKY, Helsinki.
- Östman, Jan-Ola – Fried, Mirjam 2005: The cognitive grounding of construction grammar. *Construction grammars. Cognitive grounding and theoretical extensions*, 1–13. Toim. Jan-Ola Östman – Mirjam Fried. Benjamins, Amsterdam.