

SYNONYMIA JA KÄÄNNÖSSUOMI

JOENSUUN YLIOPISTON HUMANISTISIA JULKAISUJA
UNIVERSITY OF JOENSUU PUBLICATIONS IN THE HUMANITIES
N:O 35

Jarmo Harri Jantunen

SYNONYMIA JA KÄÄNNÖSSUOMI

**Korpusnäkökulma samamerkityksisyyden
kontekstuaalisuuteen ja
käännöskielen leksikaalisiin erityispiirteisiin**

JOENSUUN YLIOPISTO
JOENSUU 2004

Julkaisija
Publisher Joensuun yliopisto
University of Joensuu

Toimittaja
Editor Marjatta Palander

Vaihdot Joensuun yliopiston kirjasto
PL 107, 80101 Joensuu
puh. (013) 251 2677
fax (013) 251 2691
e-mail: vaihdot@joensuu.fi

Exchanges Joensuu University Library
P.O. Box 107, 80101 Joensuu
tel.+ 358-13-251 2677
fax + 358-13-251 2691
e-mail: vaihdot@joensuu.fi

Myynti Joensuun yliopiston kirjasto / Julkaisujen myynti
PL 107, 80101 Joensuu
puh. (013) 251 2652, 251 4509
fax (013) 251 2691
e-mail: joepub@joensuu.fi

Sales Joensuu University Library / Publication Sales
P.O. Box 107, 80101 Joensuu
tel.+ 358-13-251 2652, 251 4509
fax + 358-13-251 2691
e-mail: joepub@joensuu.fi

Kansi
Cover Janne Kuivalainen / Joensuun yliopistopaino
Jarmo H. Jantunen

ISBN 952-458-479-4
ISSN 0781-0369

Copyright © 2004 Jarmo Harri Jantunen
Joensuun yliopistopaino
Joensuu

ALKUSANAT

Väitöskirjani juuret ulottuvat maisteriopintojeni alkutaipaleelle: vuonna 1992 sain suomen kielen proseminaarissa silloiselta yliassistentilta Marjatta Palanderilta aiheeksi *varsin*-sanan merkityksen tarkastelun suomen murteissa. *Varsin* jatkoi aiheenani aina pro graduun asti, mutta väitöskirjassani sanan tilalla on sen kolme kuomaa ja näkökulmakin on kokenut muutoksen. Näiden asteen määrittäjien asema on säilynyt tärkeänä läpi tähänastisen tutkijantaipaleeni. Pohdinnan parissa on nyt kulunut vuosia tusina, ja tässä tuotoksessani on tarjolla tämänhetkisiä aatoksiani.

Väitöskirjalleni tarjosi aiheen professori Anna Mauranen; kiitos että nostit käännösuomen tutkimuksen esille ja perustit ryhmän sen piirteitä selvittämään! Ohjauksessasi on ollut hyvä työskennellä, sillä laaja asiantuntemuksesi ja luottamuksesi on ollut jatkuvasti läsnä. Työtäni ovat ohjanneet myös professorit Ilkka Savijärvi ja Sonja Tirkkonen-Condit. Niin ikään teitä kiitän arvokkaista näkemyksistä ja neuvoista sekä pystyvistä opastuksista.

Työni esitarkastajat professorit Andrew Chesterman ja Tuomas Huumo ansaitsevat erityiskiitokset: erinomaiset korjausehdotuksenne ja kannustavat lausuntonne ovat olleet kallisarvoisia työni hioutumisessa nykyiselleen.

Olen ollut onnekas voidessani tehdä tutkimustani innostuneiden ihmisten parissa ja monen eri ryhmän jäsenenä. *Käännösuomi ja kääntämisen universaalit – tutkimus korpusaineistolla* -tutkimusryhmän jäsenille haluan osoittaa kiitokseni monista oivalluksista, joita tapaamisemme tuottivat. Erityisesti haluan kiittää tutkijakoulu Langnetin kääntämisen ja tulkkauksen tutkimuksen ohjelman jatko-opiskelijakavereita ja ohjaajia. Teidän innostuksenne on ollut tarttuvaa, ja tutkimustyönne seuraaminen on avannut monia ulottuvuuksia kääntämisen ja sen tutkimisen maailmaan. Joensuulainen suomen kielen jatko-opiskelijoiden jorinaporukka tarjosi vuorostaan erinomaisia tilaisuuksia pohdiskella väitöskirjatyöskentelyn moninaisia haasteita; kiitokset Ossi Kokolle ryhmän vetämisestä ja kaikille jorisijoille tuesta.

Väitöskirjatyöni aikana olen saanut työskennellä kahdessa eri laitoksessa: Joensuun yliopiston kansainvälisen viestinnän laitoksessa Savonlinnassa ja suomen kielen ja kulttuuritieteiden laitoksessa Joensuussa. Molemmat laitokset ovat rohkaisseet eteenpäin ja tarjonneet tutkimustyötä tukevat puitteet; talojen hallinto-, atk- ja kirjastohenkilökunnalle lausun kiitokseni saumattomasta yhteistyöstä.

Kun epäröinti ja keskustelunhalu on vallannut kesken tutkimustyön, useat ihmiset ovat tarjonneet asiantuntemuksensa käyttööni ja edistäneet työni valmistumista. Kaikkia en voi nimeltä mainita, koska joukko on suuri. Maria Vilkuna, Lisa Lena Opas-Hänninen, Antti Arppe ja Sari Eskola, teitä kuitenkin kiitän kallisarvoisista ajatuksista. Myös Mona Baker (University of Manchester), Sara Laviosa (University of Bari), John Sinclair ja Elena Tognini-Bonelli (Tuscan Word Centre), José Lambert ja muut vuoden 2001 CETRAn kouluttajat ansaitsevat kiitokset seminaareista, kursseista ja henkilökohtaisesta ohjauksesta, joilla on ollut suuri merkitys työni edistymiselle. Väitöskirjani viimeistelyvaiheessa sain arvokasta apua toimittajaksi ryhtyneeltä professori Marjatta Palanderilta; kiitos neuvoistasi – ympyrä siis sulkeutui tältä erää! Joann von Weissenbergia kiitän lämpimästi englanninkielisen tiivistelmän kieliasun tarkistamisesta.

Olen voinut keskittyä väitöskirjatyöskentelyyni täyspäiväisesti lähes koko prosessin ajan. Tämän on tehnyt mahdolliseksi kielentutkimuksen tutkijakoulun Langnetin ja Joensuun yliopiston humanistisen tiedekunnan virka tutkijakoulutettavana ja myöhemmin työskentely tiedekunnan apurahalla; viimeiset viilaukset tein suomen kielen yliassistentin virkaa hoitaessani. Lappeenrannan kaupunkia, Savonlinnan yliopistosäätiötä, Suomen kääntäjien ja tulkkien liittoa, humanistista tiedekuntaa ja projektiamme kiitän rahoituksesta, jonka turvin olen voinut opiskella Englannissa ja Italiassa.

Ystäviäni kiitän hauskoista ja lämpimistä hetkistä, jotka tarjosivat virkistystä aherruksen lomassa, ja Jarmo Tapania tuesta ja voimasta niinäkin hetkinä, jolloin oma usko ja jaksaminen on ollut koetuksella.

Joensuussa vappuna 2004

Jarmo Harri Jantunen

Abstract

Synonymity and Translated Finnish. A Corpus-based View of Contextuality of Synonymous Expressions and Lexical Features Specific to Translated language

University of Joensuu Publications in the Humanities 35.

Jarmo Harri Jantunen

Keywords: synonyms, contextual meaning, translated Finnish, translation universals, corpus-based translation studies, degree modifiers

The aim of the present study is twofold: First, it examines how synonyms can be distinguished from each other by analysing their cotext, that is, their nearest linguistic environment. Secondly, the study focuses on untypical lexical patterning in translated language, which is hypothesised to be a translation universal. This research is a component of the project *Translated Finnish and Translation Universals: A Corpus Study* led by Anna Mauranen.

Synonymous expressions can be distinguished by several criteria, such as style, register and social or geographic dialects. In the present analysis, it is hypothesised that also cotextual co-occurrence preferences and restrictions – collocation, colligation, semantic preference and semantic prosody – strongly direct the use of synonyms and thus differentiate one item from another at the syntagmatic level. Furthermore, it is postulated that also morphological preference – both paradigmatic and syntagmatic – belongs among the components of a lexical item and distinguishes synonymous expressions.

The study also states another main hypothesis according to which cotextual profiles of lexical items in translated Finnish differ from those in non-translated Finnish; it is claimed that translations into Finnish show both untypical frequencies of lexical items and untypical lexico-grammatical as well as lexico-semantic combinations. It is suggested that this tendency occurs regardless of the source language stimulus.

The above-mentioned hypotheses are tested by studying the frequencies and association patterns of three semi-grammatical words, namely the synonymous Finnish degree modifiers *hyvin*, *kovin* and *oikein* (all roughly meaning ‘very’). The data for analysis consist of the Finnish Comparable Corpus of Fiction (FCCF), which is a subset of the Corpus of Translated Finnish (CTF) compiled at Savonlinna School of Translation Studies. The FCCF is composed of three subcorpora: (a) a corpus of non-translated Finnish, (b) a multi-source-language corpus of translated Finnish (both Indo-European and Finno-Ugric source languages), and (c) a mono-source-language corpus of translated Finnish (English as the source language).

The results indicate, firstly, that the synonymous degree modifiers clearly differ in terms of their cotextual profiles. This is manifested in every category at the syntagmatic level, but especially clearly in the semantic dimension. It also turns out that the less frequent the item is, the more restricted it is to a particular cotextual patterning. The analysis of translated Finnish reveals that translated language seems to show frequencies as well as collocations and clusters that are different from those in non-translated texts. It fails, however, to show any clearly untypical cotextual profiles; semantic preferences and prosodies being the most constant in the translation process. Quantitative differences, however, are found to some extent in every category of cotextual dimensions. The data also suggest that the cotextual profiles of synonyms may differ more in translations than in non-translations. Finally, translation universals such as simplification and the impact of source language are also discussed.

SISÄLLYS

1. JOHDANTO	1
1.1. TUTKIMUSTEHTÄVÄ.....	1
1.2. KONTEKSTUALISOIVA TUTKIMUSOTE.....	4
1.3. TUTKIMUKSEN VIITEKEHYKSET.....	5
1.4. TYÖN KULKU.....	5
2. KONTEKSTUAALINEN MERKITYSYKSIKKÖ – KORPUSNÄKÖKULMA	7
2.1. FIRTHILÄINEN NÄKÖKULMA MERKITYKSEEN.....	7
2.2. TILANNEKONTEKSTISTA KOTEKSTIIN.....	10
2.3. KONTEKSTUAALINEN VALINTA LEKSIKAALISTEN YKSIKÖIDEN KÄYTTÖÄ OHJAAMASSA.....	12
2.3.1. Rajoituksia ja preferenssejä.....	12
2.3.2. Kollokaatiot: toistuvuus, tilastollisuus ja ulottuvuus.....	15
2.3.3. Kolligationaalinen valinta.....	21
2.3.4. Kontekstin semanttinen valinta.....	24
2.3.4.1. Semanttinen preferenssi.....	24
2.3.4.2. Semanttinen prosodia.....	25
2.3.5. Muotovalintoja eli morfologinen preferenssi.....	27
2.4. KOHTI INTEGROITUA KONTEKSTUAALISTA MERKITYSYKSIKKÖÄ.....	29
3. KÄÄNNÖSSUOMI OMANLAISENA KIELIMUOTONAAN	33
3.1. KÄÄNNÖSSUOMI SUOMALAISEN KULTTUURIN MONIJÄRJESTELMÄSSÄ.....	33
3.1.1. Polysysteemi- eli monijärjestelmäteorian lähtökohdat.....	34
3.1.2. Käännössuomen ja suomennosten aseman muutos.....	35
3.1.3. Monijärjestelmäteorian kritiikkiä.....	38
3.2. KORPUSPOHJAINEN KÄÄNNÖSKIELEN KUVAUS.....	39
3.2.1. Käännössuomen korpus.....	40
3.2.2. Verrannollisten käännöskorpusten hyötyjä ja rajoituksia.....	41
3.3. OLETUKSIA KÄÄNNÖSKIELEN YLEISISTÄ TENDENSSEISTÄ.....	44
3.3.1. Käännöskielen universaalihypoteesit.....	44
3.3.2. Universaalitutkimuksen käsitteiden ja tutkimusmetodologian problematiikkaa.....	50
4. SYNONYMIA: KONTEKSTUAALISUUS JA KÄÄNNÖSKIELI	53
4.1. TIETYNASTEISTA SAMUUTTA.....	53
4.2. KONTEKSTUAALINEN SYNONYMIA.....	57
4.3. SYNONYMIA JA KÄÄNNÖSKIELI.....	62
5. TUTKIMUKSEN TOTEUTTAMINEN: HYPOTEESIT, AINEISTO JA METODIT	65
5.1. TUTKIMUSHYPOTEESIT.....	65
5.2. TUTKIMUSAINEISTONA VERRANNOLLINEN KAUNOKIRJALLISUUDEN KORPUS.....	66
5.3. HAVAINTOYKSIKÖIDEN VALINTA.....	68

5.3.1. Perusteet havaintoyksiköiden valinnalle	68
5.3.2. Astemääritteet	70
5.4. TUTKIMUSMENETELMÄT	74
5.4.1. Vertailumenetelmät	74
5.4.2. Tilastolliset menetelmät	76
5.4.3. Hakutermit, kontekstin analyysi ja tarkastelualue	80
6. SYNONYIMIT <i>HYVIN, KOVIN JA OIKEIN</i> KONTEKSTUAALISINA MERKITYSYKSIKÖINÄ	85
6.1. KOLLOKATIONAALINEN VALINTA	85
6.1.1. Kollokationaalinen valinta tarkastelualueella	85
6.1.1.1. Kollokaattien tilastollinen merkitsevyys	85
6.1.1.2. Kollokaattien esiintymistaajuus	89
6.1.2. Kollokationaalinen diversiteetti	98
6.1.3. Synonyymiset <i>hyvin, kovin</i> ja <i>oikein</i> klustereissa	101
6.1.4. Synonyymien <i>hyvin, kovin</i> ja <i>oikein</i> kollokationaalinen profiili	105
6.2. KOLLIGATIONAALINEN VALINTA	106
6.2.1. Kolligationaalinen valinta tarkastelualueella	106
6.2.2. Kolligationaalinen valinta positioittain tarkastelualueella	108
6.2.3. Kolligaattien klusteroituminen	116
6.2.4. Lausepituus synonyymeja erottamassa	122
6.2.5. Synonyymien <i>hyvin, kovin</i> ja <i>oikein</i> kolligationaalinen profiili	124
6.3. MORFOLOGINEN PREFERENSSI	127
6.3.1. Paradigmaattinen morfologinen preferenssi	128
6.3.2. Kontekstuaalinen morfologinen preferenssi	129
6.4. SEMANTTINEN VALINTA	132
6.4.1. Semanttinen preferenssi	133
6.4.2. Semanttinen prosodia: positiivinen ja negatiivinen aura	140
6.5. KOKOAVA KATSAUS: MÄÄRITTEIDEN <i>HYVIN, KOVIN</i> JA <i>OIKEIN</i> KONTEKSTUAALISET PROFIILIT JA SYNONYYMISYYS	146
7. KONTEKSTUAALISET MERKITYSYKSIKÖT JA SYNONYMIA KÄÄNNÖSSUOMESSA	152
7.1. ASTEMÄÄRITTEIDEN ESIINTYMISTIHEYDÄN AINEISTOISSA	152
7.2. ASTEMÄÄRITTEIDEN KONTEKSTUAALINEN VALINTA KÄÄNNÖSSUOMESSA ALKUPERÄISSUOMEEN VERRATTUNA	155
7.2.1. <i>hyvin</i>	155
7.2.1.1. Kollokationaalinen valinta	155
7.2.1.2. Kolligationaalinen valinta	165
7.2.1.3. Morfologinen preferenssi	168
7.2.1.4. Semanttinen preferenssi ja prosodia	170
7.2.1.5. <i>hyvin</i> -määritteen kontekstuaalinen merkitysyksikkö käännössuomessa	174
7.2.2. <i>kovin</i>	176
7.2.2.1. Kollokationaalinen valinta	176
7.2.2.2. Kolligationaalinen valinta	183
7.2.2.3. Morfologinen preferenssi	185

7.2.2.4. Semanttinen preferenssi ja prosodia	186
7.2.2.5. <i>kovin</i> -määritteen kontekstuaalinen merkitysyksikkö käännössuomessa	188
7.2.3. <i>oikein</i>	190
7.2.3.1. Kollokationaalinen valinta	190
7.2.3.2. Kolligationaalinen valinta	196
7.2.3.3. Morfologinen preferenssi	200
7.2.3.4. Semanttinen preferenssi ja prosodia	200
7.2.3.5. <i>oikein</i> -määritteen kontekstuaalinen merkitysyksikkö käännössuomessa	203
7.3. SYNONYMIAN MUUTTUMINEN KÄÄNNÖSSUOMESSA	204
7.3.1. Kollokaatiot	204
7.3.2. Kolligaatiot	209
7.3.3. Paradigmaattinen morfologinen preferenssi	211
7.3.4. Semanttinen preferenssi ja prosodia	213
7.4. KOKOAVA KATSAUS: OLETUKSIA JA HAVAINTOJA KÄÄNNÖSKIELEN LEKSIKAALISISTA OMINAISUUKSISTA	215
8. LOPPUKATSAUS	220
LÄHTEET	227
LIITTEET	
Liite 1. Astemääritteiden kolligaattien osuudet AS-aineistossa.	
Liite 2. O1-kollokaattien ryhmittely semanttista preferenssiä kuvaaviin luokkiin.	
Liite 3. <i>hyvin</i> - ja <i>kovin</i> -määritteiden kollokaatit aineistoittain.	

Errata

1. JOHDANTO

Synonymia ja kääntäminen ovat molemmat kielen ilmiöitä, jotka ovat herättäneet mielenkiinnon paitsi kielentutkijoissa myös tavallisissa kielenkäyttäjissä. Harvassa lienevät ne, jotka eivät olisi joskus pohtineet yhden esineen, asian tai ilmiön useita erilaisia nimityksiä tai samaa tarkoittavien mutta muodoltaan erilaisten sanojen sopivuutta johonkin kielenkäyttötilanteeseen. Harvinaisia lienevät myös ne suomalaiset, jotka eivät olisi törmänneet suomennoksiin, ehkä miettineet niiden kieltä ja peräti olleet suomennoksista tai kielestä jotain mieltä. Tämän tutkimuksen peruslähtökohdat lienevät siis ainakin jossain määrin kaikille suomalaisille tuttuja.

1.1. Tutkimustehtävä

*Ukki, vaari ja isoisä; kaunis, viehkeä ja soma – entäpä hyvin, kovin ja oikein? Kuu-
lostavatko synonyymisilta? Eittämättä. Sekä yleissanakirjojen että synonyymisanastojen mukaan edellä mainitut sanat ovat toistensa synonyymeja eli samaa tarkoittavia ilmauksia. Ja yhtäläisyyksiä sanojen merkityksissä tottakai onkin. Mutta miten on mahdollista, että kielessä, joka yleensä pyrki taloudellisuuteen, on runsain mitoin sanoja, joilla on samanlainen merkitys? Onko sanastossamme siis joutomiehiä?*

Tuskinpa. Koska kielestä ajan mittaan karsiutuvat pois ainekset, joille ei ole käyttöä, ei voida olettaa, että sanastoonkaan pesiytyisi sellaisia sanoja, joiden merkitykset olisivat identtisiä; täysin samaa tarkoittavista sanoista osa häviäisi joka tapauksessa pois. Mikä sitten pitää synonyymiset sanat hengissä ja vieläpä elinvoimaisina? Vastaus piilee erikoistumisessa. Sanastossa on käynnissä jatkuva erikoistuminen: kieleen tulevat uudet sanat saavat erikoistehtävänsä ja -ympäristönsä, ja vanhojen tehtävä voi muuttua, osa olemassa olevista taas säilyy ennallaan. Synonyymien osalta erikoistuminen tarkoittaa vakiintumista tietynlaisiin käyttötapoihin ja -ympäristöihin. Synonyymiset sanat eivät oikeastaan olekaan aivan samamerkityksisiä, kun otamme huomioon niiden käyttöominaisuudet; tässä mielessä synonymia on aina tietynasteista lähisynonymiaa. Lähisynonyymisiksi osoittautuvat edellä olevien sarjojen sanat helposti: *isoisä* on yleiskielinen sana, *vaari* ja etenkin *ukki* esimerkiksi perheenjäsenten käyttämiä rekisterisidonnaisia hellittelynimiä. Samoin *kaunis* on 'kauneutta' tarkoittava adjektiivi yleiskielessä, *soma* ja *viehkeä* lienevät käyttöyhteyksiltään jo rajoitetumpia (vrt. esim. *Tänään on kaunis ~ ?soma ~ ?viehkeä ilma; kauniimpi ~ ?somempi ~ ?viehkeämpi sukupuoli* 'naiset'). Viimeisen sarjan sanat – pääasiassa adjektiiveja ja adverbjeja määrittävät astemääritteet – ovat jo hieman monimutkaisempia. Niistä runsaasti tuonnempana.

Yhdeksi tutkimukseni päätavoitteista olenkin asettanut synonyymisten ilmausten kontekstuaalisen erilaisuuden tarkastelun. Analyysissa keskitytään leksikaalisten yksiköiden lähimpään tekstiympäristöön, kotekstiin, ja etsitään siitä löytyviä valin-

tapiirteitä, joiden perusteella rakennetaan kullekin ilmaisulle oma kontekstuaalinen profiilinsa eli kuvaus leksikaalisen yksikön kontekstuaalisesta valinnasta. Kuvauksen avulla pyritään selvittämään synonyymisten ilmausten tyypilliset käyttöympäristöt ja tutkimaan, millaisia ovat ne lähimmässä tekstiympäristössä erottuvat kontekstuaaliset ominaisuudet, jotka voivat erottaa synonyymeiksi mainittuja ilmauksia toisistaan.

Toinen tutkimusongelmani koskee käännössuomen olemusta alkuperäissuomen rinnalla. Käännössuomen eli käännösprosessin tuloksena syntyneen suomen kielen tutkimus on ollut suhteellisen vähäistä, joskin suomen kirjakielen syntyvaiheita ja kehitystä tarkasteleva tutkimus on pohjautunut suomennosmateriaaliin luonnollisesti siitä syystä, että kirjakieli on syntynyt ja 1800-luvun loppupuolelle asti kehittynyt pitkälti käännösten varassa. Käännössuomea on siten väistämättä käytetty lukuisten kirjakieleen kohdistuneiden tutkimusten materiaalina, mutta kuvauskohteenä ei yleensä ole ollut *käännössuomi* ja sen ominaispiirteet sinänsä, vaan suomennosaineiston avulla on kuvattu kirjakielen kehitystä ja olemusta.

Käännössuomen merkitystä ja määrää alkuperäissuomen rinnalla voi punnita suomennos- ja ”alkuperäissuomalaisen” kirjallisuuden suhteen avulla. Koska suomennoskirjallisuuden osuus oli painettuun, suoraan suomeksi tuotettuun kirjallisuuteen verrattuna suurempi vielä koko 1800-luvun ja vasta 1910-luvulla jälkimmäisen osuus ylitti suomennoskirjallisuuden osuuden (Kovala 1996), oli käännössuomi vielä toissa vuosisadan taitteessakin varsin viljalti käytettyä. Käännössuomella onkin ollut merkittävä tehtävä kirjasuomen, ja myös kirjallisuuden, kehityksessä.

Käännössuomen asema on säilynyt merkittävänä nykypäiviin asti, vaikka suomennosten osuus kaikesta painetusta kirjallisuudesta on nykypäivänä selvästi pienempi kuin 1800-luvun lopussa. Suomalainen kirjallisuuden kuluttaja lukee paljon käännössuomalaista tekstiä vielä nykypäivänakin: suomennetun kaunokirjallisuuden osuus on puolitoistakertaistunut vuodesta 1985 vuoteen 1999, ja suomennosten nimekemäärä on kotimaiseen kaunokirjallisuuteen verrattuna myös lähes puolitoistakertainen (Minkkinen 2001). On myös muistettava, että suomalainen elää käännössuomen ympäröimänä kaikkina vuorokauden aikoina: suomennoksia hän tapaa niin sampoopullon kyljestä kuin muropaketin tuoteselosteistakin, sanomalehtien ja aikakauslehtien uutisista, television ja elokuvien selostus- ja ruututeksteistä ja mainosten jälkiäänityksistä, monikansallisten yritysten tuotekuvastoista ja monista muista lähteistä, joiden käännöskielisyyttä vastaanottaja ei aina edes tule huomaneeksi.

Käännöksiä käsiteltäessä kiinnittyy huomio usein siihen, että itse suomennoksessa on jotain vikaa suhteessa lähdetekstiin tai sitten suomennosten kielessä on jotain moitittavaa (käännöskriittikistä mm. Chesterman 2000). Etenkin jälkimmäiset huomiot ovat tavallisia, koska niitä pystyy tekemään perehtymättä kahden kielisysteemin väliseen vastaavuuden ongelmaan. Sen sijaan deskriptiivinen analyysi, joka ei pyri ohjaamaan ja osoittamaan, on ollut ominaista vasta vuosituhaten vaihteen tutkimukselle (ks. Jantunen ja Eskola 2002). Myös käsillä oleva tutkimus jatkaa deskriptiivistä tutkimustraditiota ja pyrkii täydentämään suomen kielen kuvaa tarjoamalla katsauksen käännössuomen sanaston ominaisuuksiin.

Miksi sitten synonyymit ovat kiinnostavia juuri kääntämisen kannalta? Synonyymien vivahteikas käyttö on yksi varioinnin ja koheesion keino, ja se kuuluu kielensä hyvin hallitsevan kielenkäyttäjän keinoihin. Sama pätee luonnollisesti kääntämiseen ja kääntäjään. Siinä missä alkuperäistekstinkin kirjoittaja, myös kääntäjä miettii keinoja tautologian vähentämiseksi ja tekstin sidoksisuuden lisäämiseksi. Synonymia on kiintoisa ilmiö myös kahden eri kulttuurin ja kielen vuorovaikutustilanteessa. Miten valita synonyymipesyestä oikea vastine lähdekielen ilmaisulle? Synonymia onkin kiinnostanut kääntämisen ja käännöskielen tutkijoita problemaattisuutensa ja luultavasti käytännönläheisyytensä vuoksi. Synonyymiset ilmaukset ovat käännösprosessin luonteen tutkimisen kannalta kiinnostavia myös siksi, että niiden avulla voidaan oivasti perehtyä muutoksiin ja variaatioon, joita kääntämisessä yleensä ja käännöskielen leksikossa mahdollisesti esiintyy. Käännöksiä ja niiden lähdetekstejä vertailevassa tutkimuksessa on havaittu esimerkiksi synonyymisten ilmausten käytön kaventuvan käännöksissä ja vain synonyymipesyeseen tutuimpien ilmausten valikoituvan lähdekielen ilmausten vastineiksi (Blum-Kulka ja Levenston 1983; Klaudy 1996). On selvää, että synonyymien kirjo on erilainen eri kielissä, ja tämä vaikuttaa luonnollisesti myös kahden eri kielen välillä toimittaessa. Synonyymipesyiden erilaisuuteen vaikuttavat niin kulttuuriset tekijät, yhteiskuntajärjestelmät kuin vaikkapa elinolosuhteet – montakohan lunta tarkoittavaa sanaa on päiväntasaajan tienoolla puhuttavissa kielissä? Suomessa niitä on runsaasti. Kääntäjänkin on siis, ainakin jossain määrin, turvaututtava yleistyksiin.

Mutta miten on laita, kun verrataan alkuperäiskieltä eli kieltä, joka on syntynyt ”spontaanisti” ilman kääntämistä, ja käännöskieltä toisiinsa? Ei siis kahta eri kieltä, jotka ovat mukana käännösprosessissa, vaan tuotoksen kieltä ja sitä vastaavaa alkuperäiskieltä. Onko synonyymien käyttö erilaista näissä kahdessa kielimuodossa? Joidenkin mielestä on. On oletettu, että synonyymipesyiden jäsenistä valikoituisi vain osa käytettäväksi käännöskielessä ja että tällainen synonyymien käytön kaventuminen yksinkertaistaisi käännöksissä käytettyä sanastoa (Kohn 1996; ks. myös Laviosa-Braithwaite 1997). Aiemmassa pienehkössä tutkimuksessani (Jantunen 2001a) en ole kuitenkaan saanut tälle vahvistusta. Entä voivatko synonyymisiä ilmauksia ympäröivät kontekstit olla käännöskielessä erilaisia kuin alkuperäiskielessä? Tutkimukseni toisena pääasiallisena tavoitteena onkin tutkia, eroavatko alkuperäissuomen ja käännössuomen leksikaalisten yksiköiden kontekstuaaliset valinnat toisistaan. Analyysissa tarkastellaan myös sitä, poikkeako yhdestä lähdekielestä käännettyistä teksteistä tavattava kontekstuaalinen valinta yleisesti käännöskielessä esiintyvistä. Käännössuomen kontekstuaalisten suhteiden kuvaus perustuu synonyymeista tehtäviin kontekstuaalisiin havaintoihin, ja molempia kielimuotoja tarkastellaan samojen synonyymisten ilmausten avulla. Näin on myös mahdollista tutkia, mitä synonyymisyydelle tapahtuu käännöskielessä. Käännössuomen osalta tutkimuksessa tarkastellaan myös niin sanottuja käännösuniversaaleja, joita on kuvattu tarkemmin luvussa 3.

1.2. Kontekstualisoiva tutkimusote

Sekä käännöskieltä että synonymiaa tarkastellaan tutkimuksessani niitä ympäröivän ja niiden kanssa vuorovaikutuksessa olevan ympäristön avulla. Käännössuomen analyysin kehyksenä on polysysteeminen lähestymistapa (Even-Zohar 1990), jota käyttäen käännössuomea tarkastellaan osana laajempaa yhteiskunnallista ja kulttuurista kehystä. Polysysteemi- eli monijärjestelmäteoria kuvaa kulttuurisia ja yhteiskunnallisia ilmiöitä muuttuvina ja toisiinsa kytkeytyneinä osajärjestelminä, jotka ovat dynaamisessa suhteessa toisiinsa. Teorian tavoitteena on nostaa tutkimuksen keskiöön yhtä lailla arvostettuja, kulttuurissa kanonisoituja järjestelmiä kuin kaanonin ulkopuolellekin kuuluvia systeemejä, jolloin näiden kahden vertailu ja vaikutussuhteiden kuvaus tulevat mahdollisiksi. Monijärjestelmän muodostavia osajärjestelmiä ei pidetä tarkkarajaisina, stabiileina eikä itsenäisinä vaan toisiinsa vaikuttavina, rajoiltaan sumeina ja muuttuvina systeemeinä. Teoria on noussut merkittävään asemaan erityisesti käännöstutkimuksessa, kun käännöskirjallisuutta ja käännöskieltä on alettu tarkastella suhteessa ympäristöön eli muihin kulttuurisiin osajärjestelmiin.

Käsillä olevassa tutkimuksessa käännössuomea ei kuvata itsenäisenä entiteettinä vaan suhteessa toiseen kielimuotoon. Käännössuomen tutkimuksessa pääpaino on suomen kielen kahden eri muodon vertailussa, jossa käännössuomi asetetaan alkuperäissuomen rinnalle. Olennaista on, että tutkimuksen tavoitteena ei ole nostaa kumpaakaan kielimuotoa toista parempaan tai tavoiteltavampaan asemaan vaan järjestelmiä kuvataan arvottomatta. Tässä yhteydessä on syytä tarkentaa, että työssä käytettävä käsite *kielimuoto* on motivoitunut ensisijaisesti tarkasteltavien kieliaineiden syntyprosessin kautta. Käsitteen käytöllä ei haluta etukäteen ennustaa kieliaineiden erilaisuutta (vrt. esim. kirja- ja puhekieli tai murteiden erot), vaikkakin erottavia piirteitä on jo havaittu (ks. esim. Eskola 2002). Sen sijaan käsitettä käytetään osoittamaan sitä, että käännöskieli syntyy toisen olemassa olevan kirjallisen tai suullisen (yleensä) vieraskielisen kieliaineksen perusteella mutta alkuperäiskieli niin sanotusti spontaanisti ilman lähdekielen läsnäoloa (ks. kuitenkin lukua 3.1.3).

Myös kuvauksessa käytettäviä kielenaineksia, synonyymeja, tarkastellaan nimenomaan niiden ympäristön kautta. Yhdeksi merkitykseltään samankaltaisiksi esitettyjä ilmauksia erottavaksi piirteeksi esitetään niiden tekstuaalisesta lähikontekstista – kontekstista – tavattavia ilmauskohtaisia preferenssejä, jotka ohjaavat synonyymien käyttöä. Kielenulkoinen konteksti jää tutkimuksessa vähemmälle huomiolle, joskaan sitäkään ei voi kokonaan synonyymeja tutkittaessa sivuuttaa, sillä synonyymien valintaan vaikuttavat niin tekstuaalisen lähikontekstin kuin tekstin- ja kielenulkoisetkin kontekstuaaliset ominaisuudet.

Synonyymien kontekstuaalisuus on oikeastaan piirre, johon kiinnitetään huomiota hyvinkin usein: yhden ilmiön erilaisia kielellisiä todentumia tarkastellaan mieluusti esimerkiksi murteittain, jolloin itse asiassa kerrotaan synonyymipesyeen eri jäsenten kontekstuaalisista preferensseistä. Samoin on laita silloin, kun lähimerkityksisten sanojen käyttöä ohjataan tekstilaji- ja rekisterikonventioihin perustuen. Kontekstuaalistavaa lähestymistapaa ei tässä mielessä voikaan pitää mitenkään uutena. Lähikontekstiin perustuva näkökulma avaa sen sijaan joitakin uusia mah-

dollisuuksia tutkia, miten ilmaukset erottuvat tekstiympäristönsä perusteella toisistaan. Työssä vaikuttaakin kaiken kaikkiaan eräänlainen kaksoiskontekstuaalisuus: synonyymeja tarkastellaan yhtäältä tekstuaalisessa mikrokontekstissa ja toisaalta laajemmissa käännessuomen ja alkuperäissuomen systeemeissä.

1.3. Tutkimuksen viitekehykset

Tutkimukseni pyrkii osaltaan vastaamaan Joensuun yliopiston kansainvälisen viestinnän laitoksessa vuosina 1997–2001 toimineen Suomen Akatemian rahoittaman *Käännessuomi ja kääntämisen universaalit – tutkimus korpusaineistolla* -projektin tavoitteisiin. Professori Anna Maurasen luotsaaman projektin päämääränä oli yhtäältä selvittää, millaisia ovat sellaiset käännessuomelle ominaiset kielelliset piirteet, jotka erottavat tämän kielimuodon alkuperäissuomesta. Toisena tavoitteena oli tuoda lisävalaistusta niin sanottujen käännessuomien eli käännessuomien lähde- ja kohdekielistä riippumatta esiintyvien yleisten ominaisuuksien tutkimukseen ja kehittää tutkimusmetodeja. Projekti on tuottanut lukuisia kotimaisia ja kansainvälisiä julkaisuja, joita en kuitenkaan käy tässä luettelemaan niiden runsaslukuisuuden vuoksi, mutta joista suuri osa on päätynyt tähänkin työhön arvokkaiksi lähteiksi. Projektissa koottu *Käännessuomen korpus* (Mauranen 1998a, 2000b) täydentää Suomessa koostettuja elektronisista kieliaineistoja tarjoamalla tutkijoiden käyttöön kirjoitetun kielen materiaalia kahdella erilaisella prosessilla tuotettuna useista tekstilajeista. Oman tutkimukseni aineisto on myös koostettu tästä korpuksesta.

Projektista valmistuneessa ensimmäisessä väitöskirjassa, Sari Eskolan *Syntetisoivat rakenteet käännessuomessa* -tutkimuksessa (2002) ja myös aiemmissa laaja-alaisesti käännessuomien tuotoksena syntyneitä suomen kieltä tarkastelleissa Puurtisen (1995) ja Vehmas-Lehdon (1989) tutkimuksissa on käännessuomi asetettu alkuperäissuomen rinnalle ja tutkimus on perustunut näiden vertailuun. Näissä kaikissa kuvaus on keskittynyt syntaksin ilmiöihin. Tutkimuksista erityisesti Eskolan työ on tavoitteiltaan ja metodologiselta lähestymistavaltaan omalle työlleni läheinen. Eskolan ja käsillä olevassa tutkimuksessa on kuitenkin myös muitakin eroja sen lisäksi, että oma työni tarkastelee käännessuomen leksikaalisia piirteitä ja Eskolan syntaktisia: siinä missä Eskolan työssä tulee vahvasti esiin nimenomaan käännessuomien tutkimuksen näkökulma käännessuomeen, oma työni luo silmäyksen käännessuomen suomen kielen tutkimuksen perspektiivistä.

1.4. Työn kulku

Tutkimukseni etenee seuraavasti. Luvussa 2 esittelen piirteitä, joista sanojen kontekstuaalinen merkityksikkö koostuu. Esityksessäni tarkastelen, mitä eri piirteillä on aiemmassa tutkimuksessa tarkoitettu ja miten niitä on käytetty merkityksen analy-

soimisessa. Katsaus painottuu englannin kielen tutkimuksessa saatuihin havaintoihin ja luotuihin käsitteenmäärittelyihin, koska korpuspohjainen merkityksentutkimus kumpuaa pitkälti englannin kielen tutkimustraditiosta. Suomen kielestä tehtyä tutkimusta käsitellään kuitenkin mahdollisuuksien mukaan. Luvussa perehdytään kontekstuaalisen merkityksentutkimukseen aloittaen J. R. Firthistä ja ulottaen katsaus nykyisin harjoitettavaan tutkimukseen; lisäksi luvussa etsitään myös uusia kontekstuaalisia ulottuvuuksia. Luku 3 esittelee pääpiirteittäin korpuspohjaista käännöskielen tutkimusta ja tarkastelee myös käännössuomen asemaa ja aseman muutosta alkuperäissuomen rinnalla. Niin ikään luku esittelee myös sellaisia käännöskielestä esitettyjä oletuksia, jotka ovat tämän työn tavoitteiden kannalta olennaisia.

Synonymia ja erityisesti synonymian kontekstuaalisuus ja samaa tarkoittavien ilmausten käyttö käännöskielessä muodostaa vuorostaan 4. luvun sisällön. Esittelen erityisesti synonyymisten ilmauksien kontekstista erottuvia piirteitä, joiden on todettu erottavan ilmausten käyttöyhteydet toisistaan. Synonyymien ja käännöskielen suhteessa keskityn pääasiassa siihen, miten synonyymisia ilmauksia on oletettu käännöskielessä käytettävän ja millaisia ovat niiden suhteesta saadut tähänastiset tutkimustulokset.

Luvussa 5 esittelen tutkimukseni tavoitteet ja hypoteesit sekä synonymian tutkimisessa käytettävien havaintoyksiköiden valintaan vaikuttavat seikat. Luku esittelee myös tutkimusaineiston ja käytettävät metodit. Luvussa selostan myös periaatteet, joiden perusteella tarkastelukonteksti on rajattu, ja vertailen erilaisia menetelmiä, joita käännöskielen tutkimuksessa on käytetty. Käsitelen myös työssä käytettäviä tilastollisia menetelmiä. Seuraavissa luvuissa esittelen aineiston analyysistä saamani tulokset. Luku 6 keskittyy synonyymisten ilmausten kontekstuaalisiin eroihin; käsittely alkaa leksikaalisista myötäesiintymistä ja päättyy kontekstin semanttisiin ominaisuuksiin. Luku 7 puolestaan keskittyy käännössuomen leksikaaliseen kuvaukseen. Kuvaus rakentuu edellisen luvun pohjalle siten, että alkuperäissuomen aineistolla tehdyn kontekstuaalisten piirteiden analyysin tuloksia verrataan käännössuomen aineistoista saatuihin tuloksiin. Luvut muodostavat siis yhdessä kokonaisuuden. Työni viimeinen luku kokoaa tutkimukseni keskeisimmät tulokset ja vertaa niitä aiempaan tutkimukseen. Samalla pohdin myös tutkimuksessa esiin tulleita rajoituksia ja nostan esiin kysymyksiä, jotka ovat vielä avoimia ja odottavat vastauksia.

2. KONTEKSTUAALINEN MERKITYSYKSIKKÖ – KORPUSNÄKÖKULMA

2.1. Firthiläinen näkökulma merkitykseen

Kontekstuaalinen semantiikka pyrkii selvittämään merkityksen muodostumista kontekstilähtöisesti ilmausta ympäröivän lähimmän tekstikontekstin tai laajemman, koko sosiaalisen ympäristön huomioon ottavan analyysin avulla. Merkityksen kontekstuaalinen luonne on ollut niin kutsutun brittiläisen koulukunnan – erityisesti J. R. Firthin, M. A. K. Hallidayn ja John Sinclairin – tutkimuskohteena (Stubbs 1996: 25). Firthiläisen kieliteorian kulmakivenä on aineistoon perustuva analyysi, se että kieltä tutkitaan todelliseen kielenkäyttöön perustuen (ks. mm. Palmer 1976; Stubbs 1996; Tognini-Bonelli 1996; Kenny 2001). Firth korostaa, että kielentutkimuksessa olisi pitädyttävä objektiivisesti havaittavissa autenttisissa tilanteissa, joissa kieltä käytetään, ja pyrittävä kuvaamaan kieltä tilannesidoksisten relaatioiden avulla. Vaatimus perustuu käsitykseen, jonka mukaan huolellisimmankaan intuitiivisen ja introspektiivisen havainnoinnin avulla ei saada tarpeeksi tietoa kielen käytön taustalla vaikuttavista prosesseista. Kontekstuaalisen lähestymistavan ja erityisesti tilannekontekstin tutkimisen avulla päästään Firthin mielestä yleistettävälle tasolle, jolloin analyysi ei nojautu yksittäisen kielenkäyttäjän mentaalipohjaisiin havaintoihin.¹ (Firth 1968 [1957]: 170, 176.)

Lähestymistavalle on ominaista merkityksen ja muodon sekä leksikon ja kielioopin yhteyden korostaminen. Muun muassa Sinclair (1966: 411) pitää kielioopin systeemin ja leksikon kuvauksia toisiaan täydentävinä tarkastelualueina, sillä tutkimalla leksikaalisten yksiköiden systemaattisia myötäesiintymätendenssejä voidaan täydentää kielen sääntöjärjestelmän ja kielen rakenteen kuvausta. Myöhemmin (1991: 6–7) hän huomauttaa, että sanan merkitys on sidoksissa sekä erilaisiin kielipillisiin että leksikaalisiin valintoihin ja että merkitys on siten vahvasti sidoksissa rakenteeseen. Sinclairin (mts. 103) mukaan kielioipissa ei yleensä ole otettu huomioon sanastoa ja merkitystä (esimerkiksi, että tietty kielipillinen skeema on mahdollisesti ominainen vain yhdelle sanan monista merkityksistä) ja vastaavasti sanaston kuvaukseen pyrkivissä sanakirjoissa on lekseemien kontekstissa esiintyvien tyypillisten kontekstipiirteiden kuvaus vajavaista.

Kontekstuaalinen merkityksentutkimus eroaa monesta muusta tavasta analysoida merkitystä, mutta erityisesti se poikkeaa konseptuaalisesta semantiikasta. Merkityksen käsitteoria pitää merkitystä käsitteenä, ideana tai mielikuvana, jonka sana syn-

¹ Firthin kannanotto korostaa eroa chomskylaiseen kielenkäyttäjän kompetenssia ja kielikäsitystä hyödyntävään kielenkuvausmalliin. Chafe (1992: 86) kuitenkin huomauttaa, että kielentutkimus on ollut ja on edelleen pohjimmiltaan paljolti introspektioon nojautuvaa. Monet jaottelut, kuten esimerkiksi morfeemien luokittelu, perustuvat Chafen mukaan juuri kielentutkijoiden introspektioon. Stubbs (1995a: 48) esittää lisäksi, että aineistoonkaan perustuvaa tutkimusta ei voi tehdä täysin ilman tutkijan intuitiota, sillä tutkimuskysymysten syntyä, tutkimusten toteuttamista ja tulosten tulkintaa ohjaavaa pitkälti tutkijan oma intuitio.

nyttää kielenkäyttäjän tajunnassa tai jonka kielenkäyttäjä yhdistää sanaan (Kangasniemi 1997: 10). Tämä taas on tutkimuksen kannalta ongelmallista, koska mentaaliset käsitteet ovat viime kädessä olemassa kielenkäyttäjän tajunnassa ja niihin ei voi päästä käsiksi kukaan muu kuin yksilö itse, minkä vuoksi käsitteet ovat hyvin subjektiivisia (Palmer 1976: 28). Kangasniemen (mts. 11) mukaan merkityksen olemus ei muutu selkeämmäksi, jos se määritellään mentaaliseksi käsitteeksi, koska silloin joudutaan kysymään, mitä tajunnassa olevat käsitteet ovat. Kuiri (2000: 9) kuitenkin huomauttaa, että konseptualismissa (kuten kontekstuaalisessakin lähestymistavassa) on olemassa myös sosiaalinen luonteensa, sillä merkitys (käsite) ei voi olla olemassa vain yksilön tajunnassa vaan se jaetaan kollektiivisesti saman kielen puhujien kesken, jotta toisten ymmärtäminen olisi ylipäättään mahdollista.

Kontekstuaalinen semantiikka ei pyri etsimään abstrakteja käsitteitä, jotka kielenkäyttäjät jakavat, vaan se keskittyy tutkimaan konkreettisia toistuvia käyttöyhteyksiä ja näin kuvaamaan niiden usuaalista merkitystä. Tutkimuksen keskiössä on erityisesti ilmausten syntagmaattisten ja paradigmaattisten ulottuvuuksien kohtauspiste eli ne säännöt tai preferenssit, joiden perusteella ilmaus valitaan muiden ilmusten joukosta (paradigmaattinen suhde) muiden ilmausten kanssa lineaariseen jatkumoon (syntagmaattinen suhde). Otetaan esimerkiksi segmentti <*pieni satuja rakastava lapsi itki tyynynsä märäksi*>, joka kuuluu laajempaan Käännösuomen korpuksesta otettuun kontekstiin:

-- ja äidin ääni kuului kauniina, taipuisana, kun hän kertoili lehteä lukevalle, tuskin kuuntelevalle isälle suvustaan ja lapsuudestaan. Äidin tarinassa **pieni satuja rakastava lapsi itki tyynynsä märäksi**, ymmärsi olleensa kopea ja turhamainen, kun oli usuttanut sisarusparvensa leikkimään Tuhkimoa.

Tästä ilmauksesta voidaan syntagmaattisessa suhteessa oleviksi ilmauksiksi (ilman tilastollista testausta) olettaa esimerkiksi tapaukset PIENI <LAPSI>, LAPSI <SATU> ja LAPSI <ITKEÄ>.² Paradigmaattisia suhteita, joista kukin sana on ilmaukseen otettu, voi taas olla monia. Kuvioon 2.1 on lueteltu muutama vaihtoehtoinen esimerkki.

Kyseisessä tapauksessa esimerkiksi *satu* on valikoitunut syntagmaattiseen suhteeseen nimenomaan *tilannekontekstin* vuoksi (käsitteestä tarkemmin seuraavassa luvussa): se, että äidin tarinassa kerrotaan lapsesta, joka rakastaa satuja eikä esimerkiksi kirjoja tai puita, on motivoinut juuri *satu*-sanan valinnan kollokaatiksi. Konseptuaalinen merkitys taas ohjaa *paradigmaattista* valintaa: fiktiivistä kertomusta merkitsevä *satu* on valikoitunut kaikista mahdollisista ilmauksista siksi, että sanalla on juuri tämä denotaatio.

² Tässä ja myöhemmin kapiteeleilla (SATU) merkitään lemmaa ja kursiiveilla tarkoitetaan vain yhtä, kyseistä sananmuotoa (*satuja*). Merkintätavassa x <y> noodi eli hakusana on kulmasulkeiden edessä (x) ja sulkeissa on ilmoitettu kollokaatti tai kollokaatit (y).



Kuvio 2.1. Syntagmaattinen ja paradigmaattinen ulottuvuus.

Kontekstuaaliseen merkitykseen³ on viitattu toisinaan käsitteellä kollokatiivinen merkitys (mm. Leech 1981; Partington 1998; Kangasniemi 1997; Kuiri 2000), jolla on tarkoitettu sanojen taipumusta esiintyä toistuvissa myötäesiintymissä toisten sanojen kanssa. Kollokatiivista merkitystä on usein selvennetty esimerkeillä, joissa on esitetty lähisynonyymipareja ja niiden kollokatiivista valintaa. Muun muassa Leech (1981: 17; ks. myös Palmer 1976: 96) antaa esimerkeiksi lähisynonyymit PRETTY ja HANDSOME ('ulkoisilta ominaisuuksilta esteettisesti miellyttävä') ja luettelee näille sanoille seuraavanlaisen kollokatiivisen jakauman:

PRETTY <GIRL, BOY, WOMAN, FLOWER, GARDEN, COLOUR, VILLAGE>
HANDSOME <BOY, MAN, CAR, VESSEL, OVERCOAT, AIRLINER, TYPEWRITER>.

Myös suomesta on löydettävissä vastaava kollokatiivinen distinktio KAUNIS- ja KOMEA-sanojen väliltä: Suomen kielen perussanakirjan (CD-Perussanakirja 1997, myöh. CD-PS) mukaan KAUNIS ja NAINEN sekä toisaalta KOMEA ja MIES muodostavat kollokaatioita⁴. On kuitenkin huomattava, että kontekstuaalisessa merkityksessä on kyse paljon laajemmasta asiasta kuin vain sanojen yhteisesiintymisestä: se käsittää niin leksikaalis-kieliopillisia kuin leksikaalis-semanttisiakin suhteita, kuten myöhemmin ilmenee. Lopuksi on vielä huomautettava, että kontekstuaalinen ja konseptuaalinen merkitysanalyysi eivät ole toisiaan poissulkevia vaan pikemminkin täydentäviä tarkastelutapoja. Siinä missä konseptuaalinen merkitysanalyysi pyrkii löytämään kielenkäyttäjille yhteisiä käsitemaailmoja, kontekstuaalinen lähtökohta pyrkii kuvaamaan todellista ja yhteistä kielenkäyttöä. Lisäksi on vielä huomattava,

³ Käsillä olevassa tutkimuksessa *kontekstuaalisella merkityksellä* tarkoitetaan ilmauksen kontekstuaalista valintaa, joka yhdessä ilmauksen kanssa muodostaa kontekstuaalisen merkitysyksikön (vrt. *extended unit of meaning*, Sinclair 1996, Stubbs 2001). Käsitteellä ei viitata tilannekohtaiseen konseptuaaliseen merkitysvariaatioon, jonka perusteella ilmauksen merkitys voi poiketa referatiivisesta merkityksestä.

⁴ KOMEA <NAINEN> ja KAUNIS <MIES> ovat myös mahdollisia kollokaatioita, mutta rikkoessaan kollokatiivista distinktiota niillä on myös konnotatiivisia merkityksiä.

että korpuslingvistinen kontekstuaalinen merkityksen tutkimukseen ei ole tullut toimeen ilman käsiteanalyysia. Erityisesti tämä on havaittavissa semanttista preferenssiä ja prosodiaa (tarkemmin luvussa 2.3.4) analysoimaan pyrkivissä tutkimuksissa, joissa leksikaalisten yksiköiden kontekstin eli lähimmän tekstiympäristön semanttisia piirteitä on kuvattu hyvinkin epämääräisillä käsitteillä, kuten 'negatiivinen' ja 'positiivinen' tai 'epämiellyttävä', 'hyvä' ja 'huono'.

2.2. Tilannekontekstista kontekstiin

Kuten edellä tuli ilmi, firthiläisen kielentutkimuksen kulmakivenä on aineistoon perustuva analyysi, se että kieltä tutkitaan todelliseen kielenkäyttöön perustuen. Seuraava, englannin fonologian käsittelyn yhdessä esiintyvä Firthin kannanotto kuvaa hyvin hänen näkemystään kontekstipohjaisen tutkimuksen tähdellisyydestä:

We must take our facts from speech sequences, verbally complete in themselves and operating in contexts of situation which are typical, recurrent, and repeatedly observable (1957: 35).

Tutkimuksen lähtökohdaksi on Firthin tavoitteiden mukaan otettava kielenkäyttötilanteeseen sidoksissa oleva kieli eli Chafen (1992: 84–88) käsitteitä soveltaen luonnollisesti esiintyvän näkyvän toiminnan havainnoiminen. Tutkimuskohteena on siis performanssi eli konkreettinen tilannesidoksinen kielenkäyttö, ja lingvistiset päätelmät on tehtävä tämän havainnoinnin perusteella. Firthin teoria kielen kontekstuaalisesta luonteesta ottaa huomioon sekä tekstinsisäisen että -ulkoisen kontekstin. Toisaalta huomio on kiinnitettävä itse tekstissä oleviin kontekstuaalisiin piirteisiin kuten syntagmaattisiin (esim. kollokointi) ja paradigmaattisiin suhteisiin (esim. synonyimia), ja toisaalta on tärkeää ottaa huomioon myös kielen ulkopuolinen konteksti eli se todellisuus, jossa kommunikointi tapahtuu. (Firth 1968 [1957]: 173.) Firth tarkastelee kieltä erottamattomana osana ihmisen sosiaalista toimintaa, johon kuuluvat niin kielelliset kuin ei-kielellisetkin kommunikointikeinot ja jota ohjaa kaikkinaisen sosiaalinen vuorovaikutus ja laajasti ottaen koko kulttuurikonteksti. Tognini-Bonelli (1996: 27) toteaaakin, että Firthin perimmäisenä tarkoituksena oli kuvata ihmistä kokonaisuutena, aktiivisena sosiaalisena toimijana.

Käsitteitä *tilannekonteksti*, *konteksti* ja *koteksti* on jo edellä tarvittu kontekstuaalisen semantiikan peruseräkkeiden selvittämisessä. Koska käsitteiden käyttö on melko kirjavaa, on paikallaan lyhyesti vielä tarkentaa, mitä näillä käsitteillä tarkoitetaan tässä tutkimuksessa. Firth (1968 [1957]: 175) kuvaa kielen ja kontekstin suhdetta käsitteellä *tilannekonteksti* (*context of situation*)⁵. Hänen mukaansa jokainen teksti kuuluu aina johonkin kielenkäyttötilanteeseen ja jokaisen ilmauksen merkitys riippuu tilanteesta käyttötarkoituksesta. Samaa painottaa myös Lyons (1968: 410), jonka mukaan sanojen merkityksen ymmärtäminen voi olla jopa mahdotonta ilman

⁵Firth on lainannut käsitteen Malinowskilta (1930), joka käytti tilannekontekstia hyväkseen tulkittessaan keräämäänsä materiaalia antropologisissa tutkimuksissa (Palmer 1976: 46–48).

niitä ympäröivää kielenkäyttöyhteyttä. Firth pitää tilannekontekstia abstraktiona, joka muodostuu toistensa kanssa vuorovaikutuksessa olevista verbaalisista ja non-verbaalisista kategorioista. Tilannekontekstissa ovat keskenään vuorovaikutuksessa kielenkäyttäjien verbaalinen ja nonverbaalinen toiminta, itse tilanne sekä viestinnän tarkoitus ja vaikutus (Firth 1968 [1957]: 175–177).

Tilannekontekstin määrittelyssä on Firthin mukaan olennaista toistuvuus:

The context of situation in the present theory is a schematic construct for application especially to typical 'repetitive events' in the social process (Firth 1968 [1957]: 176).

Toistuvuuden vaatimus tarkoittaa sitä, että tilannekontekstia ei pidä käsittää yksittäiseksi kommunikointitapahtumaksi vaan useiden rutiininomaisten kielenkäyttötilanteiden muodostamaksi abstraktioksi. Spatiotemporaaliseen tilannekontekstiin katsotaan kuuluvan myös viestin lähettäjän ja vastaanottajan yhteinen tieto siitä, mitä kielenkäyttötilanteessa on lausuttu aiemmin ja mitkä ovat tilanteelle tyypilliset konventiot. Näin ollen merkitys ei synny pelkästään itse tekstistä, vaan tekstien vuorovaikutuksesta puhuja–kuuliija-suhteessa. Kontekstuaalisen merkityksen tulkinta perustuu systemaattisen käyttäytymisen havainnointiin ja sosiaalisten prosessien tyypillisten rakenteiden analysointiin. Tilannekonteksti on siten toistuva skeemaattinen rakenne, joka tulee ilmi sosiaalisissa prosesseissa. (Firth 1957: 183, 187, 1968 [1957]: 176; Lyons 1968: 413; Mitchell 1975: 7.) Toistuvuuden vaatimukseen täytyy kuitenkin sisältyä likimääräisyyden hyväksyminen, sillä yksikään kielenkäyttötilanne ei voi toistua kahta kertaa täysin identtisenä.

Vaikka Firthin kontekstuaalista merkitysteoriaa on toisinaan pidetty liian laajana ja epämääräisenä, jotta se voisi tarjota yhtenäisen teorian merkityksestä (mm. Lyons 1966), on Firthin käsitys merkityksen kontekstuaalisesta luonteesta kantanut nykyiseen kielentutkimukseen asti, ja hänen käyttämänsä käsitteet kuten *kollokaatio* ja *kolligaatio* ovat edelleen muun muassa korpuslingvistiikan peruskäsitteistöä. Toistuvuuden vaatimus on puolestaan muodostunut yhdeksi korpuslingvistiikan perusoletuksista, kuten myöhemmin tulee ilmi kollokaatiota käsittelevässä luvussa 2.3.2.

Korpuslingvistisessä kielentutkimuksessa merkitystä on pyritty analysoimaan yleensä Firthin tilannekontekstia suppeamman *kontekstin* avulla, joskin viimeaikaisessa ilmausten semanttiseen prosodiaan liittyviä ominaisuuksia analysoivassa tutkimuksessa on tarkastelualaa laajennettu tekstikontekstista takaisin tekstinulkoiseen maailmaan (ks. lukua 2.3.4). Konteksti-käsitettä on käytetty viittaamassa hyvin monenlaisiin ympäristöihin (ks. Heikkinen 1999; Heikkinen ym. 2000). Esimerkiksi opetusministeriön asettama kieliopin sisältöä ja merkitystä pohtinut Kielioppityöryhmä rajaa mietinnössään *Kieli ja sen kieliopit* (1994) kontekstiksi sellaiset tekstinulkoiset tekijät, jotka vaikuttavat tekstin tulkintaan. Heikkinen (1999: 68) kuitenkin huomauttaa, että aina ei ole selvää, milloin teksti loppuu ja mistä konteksti alkaa, jolloin kontekstiin kuuluvien tekijöiden nimeäminen voi muodostua vaikeaksi.

Usein kontekstia pyritään jakamaan osiin määrittelemällä. Esimerkiksi Saukkonen (1984: 10) käsittää kontekstin kielenulkoiseksi kokonaisympäristöksi, jossa kieltä käytetään. Tässä ympäristössä vaikuttavat vakioituneet kielenkäyttönormit, joita ohjaa sosiaalinen ympäristö. Tässä kohdin Saukkosen näkemys on lähellä ti-

lannekontekstin käsitettä. Toisaalta kontekstilla on usein tarkoitettu ilmiön konkreettista tekstiyhteyttä, ympäröivää sanastoa, virkettä tai tätä laajempaa tekstiympäristöä. Kuitenkin muun muassa Sinclair (1991: 171–172) pitää käsitettä *koteksti* soveltuvampana kuvaamaan tutkimuksen kohteena olevaa (tarkkarajaista) kielellistä lähiympäristöä. Koteksti ei kuitenkaan käsitä pelkästään lauseyhteyttä⁶ vaan kotekstiksi voidaan laskea myös lause- ja virkerajat ylittävä teksti. Erityisesti korpustutkimuksessa analyysin alkuvaiheeseen kuuluu tarkastelukotekstin määrittäminen; tällä tarkoitan nimenomaan sitä tekstialuetta, jonka avulla ilmausten kontekstuaalisia ominaisuuksia selvitetään.

Käsillä olevassa työssä *kotekstilla* viitataan nimenomaan tutkittavan ilmauksen tekstuaaliseen lähiympäristöön, pääasiassa ennalta määrättyyn tarkastelualueeseen (ks. lukua 5.4.3). Samoin *kotekstuaalisilla* ominaisuuksilla tarkoitetaan piirteitä, jotka on löydettävissä lähimmästä tekstikontekstista. *Tilannekonteksti* on puolestaan varattu käytettäväksi silloin kun on tarpeellista viitata laajaan tekstiympäristöön tai tekstinulkoiseen maailmaan. *Kontekstia* ja *kontekstuaalisia* piirteitä käytetään vuorostaan edellisten yläkäsitteinä, jotka sisältävät erottelemattomana sekä lähimmän tekstiympäristön että tekstinulkoisen maailman ja niihin kuuluvat piirteet.

2.3. Kontekstuaalinen valinta leksikaalisten yksiköiden käyttöä ohjaamassa

2.3.1. Rajoituksia ja preferenssejä

There are many cases in texts where the independence of the choice of words is compromised, because other patterns cut across them and constrain them (Sinclair 1996: 80).

Sanojen esiintymistä syntagmaattisessa suhteessa ohjaavat monet seikat. Yksi tavallisimmin käsitellyistä on sanajärjestys, jota voi tarkastella lausekkeiden keskinäisenä järjestyksenä lauseessa ja toisaalta lausekkeen sisäisenä sanajärjestyksenä. Kielen rakenteen kuvauksessa ja kielten typologisessa luokittelussa on pyritty etsimään kielen tyypillisimpiä sana- tai lausekejärjestyksiä; suomen osalta Hakulinen ja Karlsson (1979: 311) esimerkiksi esittävät, että suomen tyypillisin lauserakenne on SVX, jossa S = subjekti, V = verbi ja X esimerkiksi objekti tai predikaatiivi. He myös luettelevat suomen kielen peruslausetyypeille ominaiset, tunnusmerkittömät sanajärjestykset, kuten esimerkiksi eksistentiaalilauseelle rakenteen AVS (mts. 301). Rakenteet siis kuvaavat sitä sanajärjestystilannetta, joka kielen erilaisissa lauseissa tyypillisimmin vallitsee, mutta jota toisaalta voidaan myös varioida (ks. Vilkuna 1996: 32–37). Jähmeämpi varioimaan on sen sijaan lausekkeen sisäinen sanajärjestys, sillä lausekkeen jäsenten paikkaa ei voi vaihtaa samaan tapaan kuin

⁶ Vrt. Hakulinen ja Ojanen (1976), s.v. *konteksti*.

lausekkeiden paikkaa lauseessa. Niinpä nominilausekkeen *erittäin kuuma päivä* jäsenet esiintyvät tässä järjestyksessä, ja muut järjestykset ovat epäkieliopillisia: **kuuma erittäin päivä, *erittäin päivä kuuma*.⁷ Sanajärjestykseen vaikuttavat myös esimerkiksi niin sanottu viskurilaki, eli pyrkimys sijoittaa kevyet rakenteet virkkeen alkuun ja raskaat loppuun, sekä myös tekstuaaliset kytkennät, joilla ilmaistaan muun muassa tarkoitteen tunnettuutta ja tekstin sidoskeinoja (Hakulinen ja Karlsson mts. 303; Vilkuna mts. 33–34; sanajärjestyksen tehtävistä ks. myös A. Hakulinen 1975).

Rakenteet, joissa sanat eivät noudata lauseelle tai lausekkeelle tyypillisiä sanajärjestyksiä, tulkitaan epäkieliopillisiksi tai epätavallisiksi *syntaktisin* perustein; Allertonin (1984: 18) termin kysymys on *syntaktisten myötäesiintymärajoitusten* rikkomisesta. Toinen Allertonin (mts. 20) mainitsema rajoitus on *semanttinen* myötäesiintymärajoitus, joka koskee sääntöjä, jotka estävät merkityspohjaisten anomalioiden tuottamisen. Niinpä esimerkiksi lause *Mansikka kävi uimassa kuusessa* on anomaalinen semanttisin tai pragmaattisin perustein, koska se rikkoo loogista kielenulkoista todellisuutta (pragmaattisista anomaliaista ks. Hakulinen ja Karlsson 1979: 52–53). Syntaktisin perustein lause ei kuitenkaan ole ongelmallinen, vertaa lausetta *Juha kävi uimassa järvessä*. Semanttisin perustein ovat ongelmallisia myös ilmaukset *naimaton ukkomies* ja *sinkku vanhapiika*. Ensimmäinen on mahdoton, koska *ukkomiehellä* tarkoitetaan naimisissa olevaa miestä; jälkimmäinen taas on ongelmallinen siksi, että sekä *sinkku* että *vanhapiika* viittaavat naimattomuuteen ja siten määräyssuhde on vähintäänkin outo toisteisuuden vuoksi. Cruse (1986: 278–279) jakaa semanttiset myötäesiintymärajoitukset edelleen kahtia: toisen ryhmän muodostavat edellä esitellyn mukaiset, loogiikkaa rikkovat anomaliat ja toisen *kollokationaaliset* rajoitukset (*collocational restrictions*), jotka eivät ole kielenulkoisesta loogisuudesta riippuvia rajoituksia vaan kielessä itsessään esiintyviä myötäesiintymärajoituksia. Cruse antaa esimerkiksi kollokationaalisesta rajoituksesta englannin *kick the bucket* (‘kuolla’) -idiomin esiintymisen vain ihmistä tarkoittavan subjektin kera; siten lause *The aspidistra* (‘tuonenkielo’) *kicked the bucket* on hänen mukaansa kollokationaalisen rajoituksen vastainen (vrt. *?Päivänkakkara potkaisi tyhjää*).

Kollokationaaliset rajoitukset jakautuvat Crusen (1986: 281) mukaan edelleen systemaattisiin (*systematic*), puolisystemaattisiin (*semi-systematic*) ja idiosynkraattisiin (*idiosyncratic*) kollokaatorajoituksiin. Systemaattisten ja puolisystemaattisten kollokaatorajoitusten erona on Crusen jaottelun mukaan se, että puolisystemaattinen sallii jossakin määrin poikkeuksia ilmauksen tyypillisimmistä myötäesiintymärajoituksista (vrt. aiemmin esimerkki *kaunis nainen* ja *komea mies* vs. *komea nainen* ja *kaunis mies*), mutta systemaattinen on yksiselitteisesti rajoittunut valinta. Jälkimmäisestä käynee esimerkiksi vaikkapa Räsänen (2002: 36) käsittelemä ilmaus *harva se* ‘melkein, lähes joka’, joka esiintyy systemaattisesti lyhyttä ja nopeasti peräkkäin toistuvaa aikaa ilmaisevien substantiivien kera: *harva se* <*tunti, aamu, iltä, päivä, yö, viikko, ?kuukausi, ?vuosi*>). Muut kuin ajanjaksoa merkitsevät kollokaatit tulkittaneen valintaan kuulumattomiksi, samoin kuin kontekstit, joissa ei ole

⁷ Vilkuna (1996: 33) huomauttaa kuitenkin, että kielenkäyttö sisältää tilanteita, joissa sanajärjestyssäantöjä rikotaan tarkoituksellisesti. Esimerkiksi *maa kallis isien* rikkoo nominilausekkeen tyypillistä rakennetta, mutta palvelee lyriikan tarkoituseriä.

toistuvuuden mahdollisuutta (vrt. *?harva se nainen/taulu/talo/metsä on kaunis*)⁸. Idiosynkraattisella kollokaatorajoituksella Cruse taas tarkoittaa myötäesiintymää, jossa ilmauksen kollokaattien määrä on rajautunut selvästi vain muutamaan ja ne voidaan esittää luettelemalla. Systemaattinen ja idiosynkraattinen rajoitus poikkeavat toisistaan pohjimmiltaan siis siinä, että systemaattinen rajoitus perustuu kollokaattien semanttisiin ominaisuuksiin, mutta idiosynkraattisen valinnan pohjana eivät ole kollokaattien semanttiset piirteet, vaan kollokaation perusta on arbitraarinen suhde.

Sanojen kontekstuaalista valintaa on kuvannut erityisesti John Sinclair (mm. 1987, 1991, 1996), jonka käsittely ei keskity kuitenkaan pelkästään sanojen keskinäisten myötäesiintymien tarkasteluun. Hänen esittämänsä ja paljon lainatut yhdistymisperiaatteet ovat vapaan valinnan (*open choice principle*) ja idiomaattisen valinnan (*idiom principle*) periaatteet, joista ensimmäinen tarkoittaa sanojen vapaata yhdistymismahdollisuutta kieliopillisuuden täytyessä. Tällaista mahdollisuutta Sinclair (1991: 110) pitää kuitenkin tekstin tuottamisen kannalta näennäisenä, sillä sanat eivät esiinny tekstissä satunnaisesti eivätkä mielivaltaisessa järjestyksessä, eikä tekstejä luoda siten vapaan valinnan perusteella. Sen sijaan hän katsoo tekstien rakentuvan idiomaattisen valinnan mukaan.

Vaikka Sinclair (1991: 110) myöntääkin, että kielen ulkopuolisessa maailmassa yhdessä tavattavat asiat esiintyvät myös kielessä yhdessä, hän huomauttaa, että kieli sisältää paljon valintoja, jotka eivät ole riippuvaisia kielenulkoisesta todellisuudesta. Niinpä hän esittääkin, että kielenkäyttäjällä on käytettävissä suuri määrä ennakolta valmiita konstruktioita, jotka ohjaavat yksittäisten sanojen käyttöä ja tekstin rakentumista. Esimerkkinä idiomaattisesta valinnasta Sinclair mainitsee kollokaatiot, ja Kenny (2001: 100) tarkentaa, että nimenomaan *tavalliset* kollokaatiot ovat osoitus idiomaattisen valinnan toteutumisesta, mutta epätavalliset kollokaatiot taas ovat vapaan valinnan tulosta. Tässä yhteydessä on syytä huomauttaa, että Sinclairin idiomaattisen valinnan perustana ei ole yhdessä esiintyvien sanojen merkitys, vaan se vastaa pikemminkin Crusen idiosynkraattista kollokaatorajoitusta. Tämä on nähtävissä Sinclairin (mm. 1996: 82) esimerkeistä, joista yksi juontaa juurensa Hallidayn esittämään distinktion englannin STRONG- ja POWERFUL-sanojen (molempien peruserkitys 'voimakas') välillä: Hallidayn (1966: 150) mukaan kollokaatiot POWERFUL <TEA> ja STRONG <CAR> ovat epätavallisia ja rikkovat näiden lähisyronymien syntagmaattista yhdistymistä ohjaavia sääntöjä. Sanojen pitäisi olla merkityksensä perusteella keskenään vaihdettavia, mutta niiden kollokationaalinen valinta ohjaa erilaisiin syntagmaoihin (STRONG <TEA> ja POWERFUL <CAR>).⁹

⁸ Koko Käännösuomen korpuksessa *harva se* kollokoi selvästi korkeintaan viikon mittaista ajanjaksoa ilmoittavan sanan kera: *harva se* <päivä 8; yö, viikko 2; aamu, ilta, tunti 1>.

⁹ Myös generatiivisessa lingvistiikassa on käsitelty valintarajoituksia, kun on pyritty selittämään, miksi semanttisesti järjettömiä lauseita ei voi lukea kieliopillisesti hyväksyttävien lauseiden joukkoon (ks. Palmer 1976: 100–101; Lyons 1981: 147–).

Edellä on useasti puhuttu cruselaisittain kollokaatio- ja valintarajoituksista. Sanojen yhdistymistä toisten sanojen kanssa on pidettävä kuitenkin jatkumona, jonka toisessa päässä on hypoteettinen sanojen täysin vapaa valinta ja toisessa päässä täysin rajoittunut myötäesiintyminen. Näiden ääripäiden väliin jää laaja alue, jossa sanojen yhdistymistä ohjaavat tiukkoja rajoituksia lievemmat *myötäesiintymäpreferenssit*, jolloin variaatio on jossain määrin mahdollista. Onkin tarpeen erottaa selkeät valintarajoitukset tendensseistä tai preferensseistä, sillä tyypillisimmin kyse on juuri (vahvasta) preferenssistä johonkin kontekstuaaliseen piirteeseen sen sijaan että kyse olisi selkeästä joko–tai-suhteesta; tämä käy ilmi lukuisista leksikaalisten yksiköiden kontekstuaalisten valintapiirteiden analyyseistä (mm. Sinclair 1991, 1996, 1998, 2003; Stubbs 1995b, 2001a). Sanoilla voi siis olla hyvinkin laaja kirjo kollokaatteja (tai muita kontekstuaalisia ominaisuuksia) ympäristössään, mutta siitä huolimatta ne voivat osoittaa selvää preferointia joihinkin tai jonkin tyyppiisiin sanoihin. Tällöin on puhuttava ennemminkin kontekstuaalisesta preferenssistä kuin rajoituksesta. Omassa tutkimuksessani kontekstuaalisen valinnan analyysi perustuu vahvasti aineistoon pohjautuvaan havainnointiin ja aineistosta erottuviin säännönmukaisuuksiin noodin (tarkasteltavana olevan lekseemin eli hakusanan, vrt. engl. *node*) ja kontekstuaalisten muuttujien välillä; analyysissä on tarvittaessa erotettu toisistaan valintarajoitukset (mikä on *mahdollista*) ja -preferenssit (mikä on *tyypillistä*).

Kollokationaalinen valinta on kuitenkin vain yksi sanojen syntagmaattisia kytkeäjä ohjaavista tekijöistä, jotka muodostavat ilmauksen kontekstuaalisen merkityksikön. Muita tähän mennessä kirjallisuudessa käsiteltyjä kontekstuaalisia valintapiirteitä ovat kolligationaalinen valinta ja kontekstin semanttiset ominaisuudet – ja mahdollisesti myös morfologinen valinta. Erilaisia kontekstuaalisia valintapiirteitä kuvataan yksityiskohtaisemmin seuraavissa luvuissa.

2.3.2. Kollokaatiot: toistuvuus, tilastollisuus ja ulottuvuus

Vaikka termi kollokaatio on esitelty jo 1950-luvulla (Firth 1957, 1968 [1957]), sitä koskevia teoreettisia analyyseja ja kollokaatioita hyödyntäviä kielenkuvauksia on ilmestynyt vain muutamia ennen 80- ja 90-lukuja, jolloin tutkimus lisääntyi eritoten suurten korpusten suomien mahdollisuuksien myötä. Nykyisin kollokaatioanalyysit ovat korpuksiin perustuvan kontekstuaalisen kielenkuvauksen keskeisintä aluetta, ja tutkimusta on harjoitettu erityisen paljon etenkin leksikografiassa. (Sinclair 1987: 319; Hoey 1997: 4; Clear 1993: 271).¹⁰ Kollokaation määrittely on vaihdellut jossain määrin tutkijoittain ja tutkimustavoitteittain siitä kun J. R. Firth kirjoitti artikkelissaan vuonna 1957, että sanat tunnustetaan seurastaan: *You shall know a word by*

¹⁰ Pohja nykyisin tehtäville kollokaatioanalyyseille on luotu kuitenkin jo ennen suurten, niin sanottujen toisen sukupolven korpusten aikaa hyvinkin pienillä aineistoilla. Esimerkiksi Jones ja Sinclair (1974) käyttivät kollokaatioanalyyseissään vajaan 150 000 saneen aineistoa ja Berry-Rogghen (1973) optimaalisen tarkastelualan määrittelyä koskeva tutkimus perustui vain hieman yli 70 000 saneen materiaaliin.

the company it keeps! (Firth 1968 [1957]: 179). Kuten edellä tuli ilmi, yksi tärkeä osa sanojen merkityksestä on niiden syntagmaattinen leksikaalinen myötäesiintyminen toisten sanojen kanssa, joka ilmenee kontekstuaalisena leksikaalisena valintana. Perusoletus syntagmaattisesta suhteesta (Firth 1957: 196) onkin myöhemmin esitettyjä määritelmiä yhdistävä piirre. Tämän lisäksi kollokaatiota koskevissa määritelmissä toistuvat useasti muutamat muutkin piirteet, joilla ilmiötä on haluttu rajata. Käsittelen seuraavassa kolmea keskeistä määritelmissä usein esille tulevaa ominaisuutta, jotka ovat myös olennaisia käsillä olevan työn metodologian kannalta. Näitä ovat *toistuvuus*, *tilastollinen merkitsevyys* ja *ulottuvuus*. Esitys ei pyri olemaan kollokaatio-käsitteen ja -tutkimuksen osalta tyhjentävä; aiheesta on olemassa laajoja yleisesityksiä (mm. Sinclair 1991; Biber ym. 1998; Partington 1998) ja sitä on käsitelty perusteellisesti myös deskriptiivisissä analyyseissa (esim. Kenny 2001).

Toistuvuus. Kollokaationa pidetään tyypillisesti sellaista sanojen yhteisesiintymää, joka on toistuva. Jo Firth on esittänyt toistuvuuden olennaisena tilannekontekstia pienempien, kontekstuaalisten piirteiden ominaisuutena: hän kuvasi leksikon syntagmaattisia suhteita nimenomaan *tyypillisten* kollokaatioiden (*habitual collocation*) avulla (1957: 194–; ks. myös 1968 [1957]: 179–181). Määritelmissä on käytetty edelleen Firthiin nojautuen ilmausta *habitual co-occurrence* (Stubbs 1995a, 1995b, 1996) sekä ilmauksia *regularly co-occur* (Halliday ja Hasan 1976), *occur repeatedly* (Carter 1987), *recurrent sequences/co-occurrence* (Kjellmer 1984; Clear 1993; Smadja 1993) ja *frequent co-occurrence* (Sinclair 1996; Stubbs 2001a). Näillä kaikilla ilmaistaan sitä, että syntagmaattisessa suhteessa olevien sanojen yhdessä esiintymisen on oltava taajaa, jotta sitä voidaan käsitellä kollokaationa. Kielessä havaittava kollokaatioiden toistuvuus synnyttää kollokationaalisen normin (Kjellmer 1991: 123; Partington 1998: 17, 121), jota voidaan kuitenkin rikkoa esimerkiksi luovaa ja innovatiivista ilmaisua (Kjellmer 1991: 123; Beaugrande 1999: 247; kääntämisen osalta Kenny 2001) tai huumoria (Partington 1998: 121) tavoiteltaessa. Toisinaan kollokaatio-käsitettä käytetään kattamaan kuitenkin myös harvoin esiintyvät tapaukset. Clear (1993: 277) huomauttaakin, että kollokaatioina on voitu pitää sellaisiakin yhteisesiintymiä, joita tavataan aineistossa vain kerran. Hänen mukaansa kertaesiintymiin liittyy kuitenkin metodologinen ongelma: yhden korpuksen perusteella tai intuitioon nojautuen voi olla vaikea tietää, esiintyykö kerran löydetty myötäesiintymä jossakin toisessa, ehkä laajemmassa korpuksessa vai onko kyse todellisesta yksittäisestä, muualla esiintymättömästä tapauksesta. Jälkimmäiset tapauksetkin voivat olla mielenkiintoisia juuri niiden epätavallisuuden vuoksi (Sinclair 1991: 170): (tietyissä tekstilajissa) epätavalliset kollokaatiot ovat epäkonventionaalisia ja siten niillä on uutuusarvo.¹¹

¹¹ Uutuusarvo on tärkeä mm. mainosteksteissä. Mainostekstien epätavallisesta kollokaatiosta käy esimerkiksi kollokaatio *varsinaisen sydämellisesti* Helsingin Sanomien etusivulla olleessa mainoksessa, jossa kerrottiin Varsinais-Suomen esittäytymisestä Helsingissä: *Varsinais-Suomi toivottaa kaikki tervetulleiksi Senaatintorille – varsinaisen sydämellisesti* (HS 13.6.2002). Kollokaation teho perustuu tässä juuri epätavallisuuteen ja leikkittelyyn: *varsinaista* ei juurikaan käytetä astemääritteenä – toisin kuin morfologisesti läheistä *varsinmääritettä* (ks. Jantunen 1995) – ja sillä halutaan tässä astemäärityksen lisäksi viitata myös *Varsinais-Suomi*-nimen määriteosaan.

Tilastollinen merkitsevyys. Mitä taajemmin kollokaatti esiintyy, sitä merkittävämpänä sitä yleensä pidetään analyysin kannalta. Usein kuitenkin absoluuttisten tai suhteellisten frekvenssien ei ole katsottu riittävän kuvaamaan kollokaation merkitävyyttä, vaan leksikaalisia myötäesiintymiä on tarkasteltu myös tilastollisen merkitsevyyden kannalta. Tällöin aineistossa esiintyvät leksikaaliset kombinaatiot joudutaan yleensä jakamaan kolmii: a) kerran tai harvoin esiintyviin kollokaatioihin, jotka eivät ole tutkimuksen kannalta tähdellisiä, b) kollokaatioihin, jotka ylittävät tietyn, tutkimuksen kannalta relevantin esiintymistajuuden, ja c) kollokaatioihin, jotka tulkitaan merkitseviksi tilastollisten menetelmien avulla. Niinpä esimerkiksi Berry-Rogghe (1973) on jättänyt huomiotta vain kerran esiintyvät kollokaatiot, kerännyt aineistosta vähintään kaksi kertaa esiintyvät kollokaatiot ja laskenut näiden tilastolliset merkitsevyydet¹². Myös Jones ja Sinclair (1974) tuovat kollokaatiokäsitteen rinnalle *tilastollisesti* merkitsevän kollokaation käsitteen, ja analyysissa on päädytty tarkastelemaan vain tilastollisten testien merkitseviksi osoittamia kombinaatioita.

Tilastollisen testauksen tarkoituksena on löytää ne leksikaaliset kombinaatiot, joiden jäsenet esiintyvät yhdessä niin useasti, että kyse ei voi olla sattumasta (Berry-Rogghe 1973: 103). Kun sanakombinaation tilastollista merkitsevyyttä lasketaan, otetaan huomioon sanayhdistelmän jäsenten eli noodin ja sen kollokaattien todelliset frekvenssit jollakin tarkastelualueella, odotetut frekvenssit tällä alueella sekä tarkastelualueen ja aineiston koko (Barnbrook 1996: 87–106). Tämä pätee yleisimmin käytettyihin kollokaatioita kuvaaviin tilastollisiin testeihin, joita ovat *t*- ja *z*-testit sekä yhteisinformaatiotesti (Mutual Information) (mts. 94; ks. myös Biber ym. 1998: 265–268; Stubbs 1995a).¹³ Tilastollisten testien käyttö ei ole kuitenkaan kieltä analysoitaessa ongelmatonta. Testit näet olettavat, että sanat esiintyisivät tekstissä satunnaisesti, mikä on ristiriidassa kielen todellisuuden kanssa (Stubbs 1995a: 31). Kieli ei ole sattumanvaraisesti rakentunutta – ei siis myöskään sanojen asema teksteissä: niiden asemaa määräävät muun muassa syntaktiset suhteet kuten sanajärjestys, lausekkeiden keskinäinen järjestys ja niiden sisältämien sanojen järjestys, lausetyyppi ja tekstin informaatorakenne. Lisäksi sanojen esiintymistä rajoittavat myös esimerkiksi tekstin rakenteeseen ja tekstityyppiin liittyvät seikat. Tästä huolimatta tilastollisia menetelmiä käytetään työkaluina leksikaalisen sidoksisuuden analysoimisessa: menetelmillä on pystytty selvästi osoittamaan, että sanoilla on niille tyypillisiä kollokationaalisia preferenssejä, jotka eivät ole suoraan riippuvaisia edellä mainituista tekijöistä. Barnbrookin (mts. 93) mukaan satunnaiseen esiintymiseen perustuvan laskennan etuna on se, että testit eivät perustu millekään ennako-odotukselle sanojen kombinaatiomahdollisuuksista; näin ollen tekstin sanoja käsitellään neutraalisti ilman ennalta määrättyjä yhdistymismahdollisuusodotuksia. Tilastollista laskentaa on hyödynnetty muun muassa monissa leksikografi-

¹² Sinclair (1996: 81) pitää kollokaation toistumista vähintään kahdesti eräänlaisena merkitsevyyden esitestinä, sillä on epätodennäköistä, että täsmälleen sama ilmaus toistuisi sattumalta.

¹³ Kollokaatioiden laskennassa on käytetty runsaasti erilaisia testejä, joista löytyy kuvaus Oakesilta (1998).

sisä tutkimuksissa (esim. Church ja Hanks 1990; Church ym. 1991; Clear 1993) ja konekäntämisessä (ks. Church ja Mercer 1993).

Tilastolliseen analyysiin perustuvia kollokaatioita tarkasteltaessa on huomattava, että testien avulla merkitseviksi osoittautuvat kollokaatiot eivät ole välttämättä semanttiseen myötäesiintymätendenssiin perustuvia. Niinpä esimerkiksi tilastollisen laskennan perusteella merkitseviksi osoittautuneilla kollokaatioilla kuten HOUSE <THIS, OPPOSITE> (Berry-Rogghe 1973) tai TÄRKEÄ <TEKIJÄ> (Jantunen 2001b) ei ole samanlaista suhdetta kuin Karlssonin mainitsemilla merkityspohjaiseen myötäesiintymätendenssiin perustuvilla kollokaatioilla LEHMÄ <AMMUA> ja HEVONEN <HIRNUA>, joissa erityisesti verbi laukaisee valinnan (Karlsson 1999: 232). Smadja (1993: 150) kritisoikin Churchin ja Hanksin (1990) analyysia, jossa kollokaatioiksi osoittautuivat tilastollisten menetelmien perusteella muun muassa kollokaatiot *nurses* <*doctors*> ja *doctors* <*hospitals*>. Smadja (1991: 175) nimittää tällaisia kombinaatioita konseptuaalisiksi kollokaatioiksi, eli hän tulkitsee ne sanayhdistelmiksi, jotka esiintyvät taajaan yhdessä merkityssuhteidensa vuoksi. On kuitenkin huomattava, että myös semanttisesti toisilleen läheiset sanat *voivat olla* – mutta eivät *välttämättä ole* – toistensa kollokaatteja tilastollisinkin perustein; tämänhän osoittaa itse asiassa jo Churchin ja Hanksin tutkimuskin, joka perustuu nimenomaan tilastollisiin menetelmiin.¹⁴

Ulottuvuus. Kollokaation ulottuvuudella tarkoitan neljää eri seikkaa, jotka liittyvät osittain toisiinsa ja joita joudutaan pohtimaan ennen analyysin aloittamista. Näitä ovat a) kollokationaaliseen suhteeseen kuuluvien sanojen lukumäärä, b) noodin ja sen kollokaattien välimatka toisistaan, c) noodin ja sen kollokaattien kieliopillinen suhde ja d) tarkastelualueen symmetrisyys. Ensinnäkin tulkinta kollokaatioon kuuluvien sanojen lukumäärästä vaihtelee tutkijoittain: mm. Jones ja Sinclair (1974), Sinclair (1987), Hoey (1997), Biber ja Conrad (1999) sekä Steward (2000) laskevat kollokaatioksi vain kahden sanan myötäesiintymät, Kjellmer (1984) ja Smadja (1993) ovat vuorostaan pitäneet kollokaatioina sekä kaksi- että useampisanaisia kombinaatioita.

Leksikaalisia kombinaatioita kuvaavissa tarkasteluissa on kuitenkin usein ollut tarkoituksenmukaista erottaa kaksi- ja useampisanaiset lekseemikombinaatiot toisistaan. Kombinaatiot eroavat toisistaan ainakin kahdella tavalla: Ensinnäkin kollokaation jäsenten välillä voi olla sana tai useampikin (esimerkiksi LUKEA- ja KIRJA-sanojen välillä lauseessa *Luettuaan eilen illalla päivällisen jälkeen kirjaa Maija rupesi nukkumaan*), mutta useampisanaisilla lekseemikombinaatioilla käsitetään taas yleensä vain peräkkäin esiintyvistä sanoista koostuvia ketjuja. Toiseksi yhdistelmät eroavat toisistaan merkitsevyydeltään. Kjellmerin (1984: 167) mukaan on nimittäin erittäin epätodennäköistä, että useampisanaiset kombinaatiot esiintyisivät sattumalta useammin kuin kerran, ja lisäksi (toistuvan) myötäesiintymän merkitsevyys kasvaa sitä mukaa, mitä pitempi se on. Kolmi- tai useampisanaiset yhdistelmät ovat siten sekä kombinatoriselta rakenteeltaan että myös merkitsevyydeltään kak-

¹⁴ Myös Clearin (1993: 280) analyysi osoittaa, että tilastollisen testin perusteella merkitseviksi kollokaateiksi voivat osoittautua sekä semanttisin perustein yhteenkuuluvat kombinaatiot (TASTE <SALT>) että sellaiset, joiden myötäesiintyminen muodostaa kieliopillisen suhteen (TASTE <for>: *develop a taste for scifi novels*).

sisanaisista erottuvia. Useampisanaiset yhdistelmät onkin usein erotettu kollokaatioista: esimerkiksi Mauranen, Biber ja Conrad, Kenny sekä Nelson erottavat kollokaatiot *monisanaisista ketjuista* (*multi-word strings*, Mauranen 2000a, 2000b: 120), *leksikaalisista nipuista* (*lexical bundles*, Biber ja Conrad 1999), *monisanaisista yksiköistä* (*multi-word items* Nelson 2000) ja *klustereista* (*clusters*, Kenny 2000: 99; 2001), joilla kaikilla tarkoitetaan kahta sanaa pitempiä myötäesiintymäkokonaisuuksia.

Myötäesiintymätutkimuksessa lähtökohtana on pidetty sitä, että sanarypään lekseemit esiintyvät toistuvasti toistensa yhteydessä mutta niiden ei tarvitse muodostaa syntaktisesti tai semanttisesti yhtenäistä rakennetta:

Lexical bundles are defined as the most frequent recurring lexical sequences; however, they are usually *not* complete structural units, and usually not fixed expressions. Lexical bundles should be distinguished from both idioms and collocations. (Biber ja Conrad 1999: 183.)

Niinpä esimerkiksi Biber ja Conrad (mts. 186–187) luettelevat akateemisista teksteistä löydettyiksi klustereiksi muun muassa kombinaatiot *<one of the most>*, *<percent of the>* ja *<the way in which>*, jotka eivät ole syntaktisesti motivoituneita konstruktioita, ja samoin toimii myös Kenny (2001: 44) analysoidessaan käännosteksteistä tavattuja klustereita: *<our Edgar says>*, *<we say and>*, *<for a long>*, *<so as to>*. Idiomeista klusterit puolestaan erottuvat siten, että idiomit ovat kokonaisia konstruktioita ja ne voidaan toisinaan korvata yksisanaisella ilmauksella (*kick the bucket ~ die*, vrt. *heittää veivinsä ~ kuolla*) toisin kuin klusterit, jotka ovat Biberin ja Conradin mukaan idiomeja tavallisempia kielen rakenteita, joskaan eivät yhtä selvästi erottuvia (mts. 183–184).

Aiemmin mainituissa kollokaatioita käsittelevissä tutkimuksissa on kollokaation katsottu muodostuvan sanoista, jotka ovat lyhyen välimatkan päässä toisistaan. Usein myös halutaan rajata tarkasti, millä alueella olevien sanojen voidaan katsoa kollokoivan keskenään. Kollokaatioanalyysissä ensimmäisiä määrittämiä onkin tarkastelun alueen eli kollokaatiiovälin määrittäminen. Alueen määrittäminen on tärkeää, sillä tarkastelun alueen rajaaminen vaikuttaa merkittävästi tutkimustuloksiin. Mason (1997: 363) huomauttaa, että suuri tarkastelun alue voi lisätä yhtäältä tarkastelussa esiin tulevien kollokaatioiden määrää ja toisaalta yhden kollokaation frekvenssiä, mikä puolestaan vaikuttaa merkitsevyyteen. Stubbsin (1995a: 32–33) mukaan tutkimuskirjallisuudesta ei löydy kuitenkaan yhteisymmärrystä siitä, mikä on sopiva havaintoalue lekseemien kontekstuaalisia valintoja tarkasteltaessa. Hän esittää yleisimmin käytetyiksi havaintoalueiksi kollokaatiiovälejä 2:2 ja 3:3, mikä tarkoittaa kahta tai kolmea sanaa hakusanan molemmin puolin. Berry-Rogghe (1973) sekä Jones ja Sinclair (1974) vuorostaan määrittelevät optimaaliseksi havaintoalueeksi neljä sanaa hakusanan molemmin puolin (4:4). Tähän tulokseen ovat jälkimmäiset päätyneet tutkimalla positioiden 1–10 kollokaatioiden lukumäärää: heidän tuloksensa mukaan neljättä positiota kauempana olevat lekseemit kollokoivat noodin kanssa heikosti. Optimaalisen tarkastelun alueen määrittäminen vaihtelee siis tutkimuksittain.

Tarkastelualueen rajaamiseen liittyy saneissa laskettavan välimatkan lisäksi myös kieliopillisten rakenteiden huomioon ottaminen tarkastelualueella. Esimerkiksi Jones ja Sinclair sekä myöhemmin Stubbs ovat määritelleet tarkastelualueen siten, että siinä ei oteta kieliopillisia rakenteita huomioon:

-- a span has been defined by specifying a standard number of orthographic words, *disregarding the grammatical structures of which they form a part* (Jones ja Sinclair 1974: 21; korostus JJ).

Collocation -- is a purely lexical relation between words in linear sequence, *irrespective of any intervening syntactic boundaries* (Stubbs 1995b: 246; korostus JJ).

Määritelmien mukaan kollokaatioiksi lasketaan kaikki tarkastelualueen sanat piittaamatta niistä syntaktisista rakenteista, joissa lekseemejä on käytetty. Kjellmer taas korostaa kollokaation määrittelyssä kieliopillisuutta ja samanmuotoisuutta:

A collocation is a sequence of words that occurs more than once in identical form -- and which is *grammatically well-structured* (Kjellmer 1987: 133; ks. myös 1990, 1991; korostus JJ).

Käytännössä tämä tarkoittaa sitä, että Kjellmerin määritelmän mukaan kollokaatioita eivät voisi olla samaan semanttiseen kenttään kuuluvat Smadjan sekä Churchin ja Hanksin esittämät kollokaatiot LAW <COURT> ja *nurses <doctors>*, jotka eivät muodosta kieliopillista rakennetta vaan ovat kollokaatioita tilastollisin ja semanttisin perustein. Kjellmerin vaatimus kieliopillisesta rakenteesta sulkee siten pois kollokaatiot, joiden jäsenet eivät ole vierekkäisiä.

Kollokationaalaisia suhteita tarkasteltaessa joudutaan kiinnittämään huomiota myös lause- ja virkerajan merkitykseen. Esimerkiksi Smadjan (1993) analyysissa virkerajan ylittäviä kollokaatteja ei ole laskettu mukaan, mikä johtuu siitä, että tarkastelun kohteena ovat yhden virkkeen sisällä esiintyvissä syntaktisissa suhteissa, kuten verbin ja objektin muodostamassa suhteessa, olevat lekseemit. Clear (1993: 276) pitääkin epätodennäköisenä, että noodit kollokoisivat merkitsevästi toisessa virkkeessä olevien lekseemien kanssa. Edellä esitetty Kjellmerin määritelmä sulkee puolestaan automaattisesti lause- tai virkerajan ylittävät sanayhdistelmät pois. Halliday (1966: 151) ja Berry-Rogghe (1973: 108–109) kuitenkin huomauttavat, että kaikki merkitsevät kollokaatit eivät välttämättä löydy virkerajan sisältä, vaan virkerajat ylittävien viittaussuhteiden vuoksi merkitseviä kollokaatteja voi esiintyä myös edellä tai jäljessä olevista lauseista. Tästä antaa englanninkielisen esimerkin Mitchell (1971: 53): sanat JOB ja APPLY kollokoivat kiistatta keskenään, vaikka ovatkin kahdessa eri virkkeessä *He didn't want the job. I don't think he even applied* (ks. myös Halliday 1966: 151; Berry-Rogghe 1973: 109). Tällainen virke- ja lauserajat ylittävä kollokointi on yksi tekstien leksikaalisen koheesion ilmentymistä (Halliday ja Hasan 1976: 284–288); Hallidayn ja Hasanin termin kyse on kolloka-

tionaalisesta koheesiosta, jolle on tyypillistä riippumattomuus kieliopillisista rakenteista (mts. 286).¹⁵

Neljäs tarkastelualueeseen liittyvä ominaisuus on alueen symmetrisyys. Useissa kollokaatioanalyyseissa tarkastelualueeksi on määritelty symmetrinen, hakusanan molemmilta puolilta samankokoinen alue, kuten neljä sanaa noodin molemmin puolin. Tarkastelualueen symmetrisyyteen suhtautuu kuitenkin kriittisesti Mason (1997, 2000), joka osoittaa, että sanojen suhde kontekstiin ei ole välttämättä symmetristä. Symmetrinen alue jättää huomiotta sen, että eri sanoilla voi olla erilainen vaikutus leksikaaliseen kontekstiinsa, jolloin sanojen optimaalinen tarkastelualue voi olla erilainen. Tätä tutkiakseen Mason on verrannut useiden sanojen leksikaalista gravitaatiota (*lexical gravity*) laskemalla usean position (15:15) lemma-sane- ja sananmuoto-sanesuhteet¹⁶ eli toisaalta positiossa esiintyvien lekseemien ja toisaalta yksittäisten sananmuototyyppien suhteen kaikkiin position saneisiin. Laskennan avulla Mason pystyy osoittamaan, että hakusanan suhde ympäristöönsä ei ole symmetristä, vaan leksikaalinen variaatio on erilaista eri puolilla sanaa. Lisäksi analyysi osoittaa, että eri sanaluokkaan kuuluvilla sanoilla ja jopa eri sananmuodoilla voi olla erilainen leksikaalinen vetovoima. Myös Sinclair (1991, 1996) ja Stubbs (2001a: 29) ovat tuoneet esiin mahdollisuuden tarkastella lekseemejä balansoimattomalla kollokaatiovälillä. Myös useassa muussa tutkimuksessa tarkastelualue on toispuoleinen: esimerkiksi Bäcklundin (1973) ja Paradisin (1997) intensiteettiadverbeja koskeissa tutkielmissa keskitytään vain oikeanpuoleiseen alueeseen, ja siinäkin vain vieruskollokaatteihin eli adverbien pääsanoihin. Toispuoleisuuden valinta ei näissä tutkimuksissa kuitenkaan perustu leksikaaliseen gravitaatioon, vaan niissä on haluttu keskittyä yksinomaan pääsanan ja sen määritteiden kollokointiin. Käsillä olevan työn kannalta olennaiset toistuvuuteen, tilastolliseen merkitsevyyteen ja ulottuvuuteen liittyvät periaatteet on selvitetty yksityiskohtaisesti luvussa 5.

2.3.3. Kolligatiivinen valinta

Niin Halliday (1966: 159), Greenbaum (1970: 10–11), Mitchell (1971: 48) kuin Hoeykin (1997: 7) korostavat leksikaalis-kieliopillisten suhteiden kuvaamista kollokaatioiden rinnalla lekseemien kontekstuaalisen valinnan analysoinnissa, sillä lekseemit voivat esiintyä niin vakiintuneissa kieliopillisissa kuin leksikaalisissakin rakenteissa. Carter (1987: 56) esittääkin sanan merkityksen muodostuvan erilaisten syntagmaattisten suhteiden verkostosta, ja tämä verkosto on kuvattava mahdolli-

¹⁵ Hallidayn ja Hasanin antamista esimerkeistä ilmenee, että he käsittävät kollokatiivisen koheesion sellaisten sanojen synnyttämäksi, joilla on merkityspohjainen taipumus esiintyä yhdessä. Esimerkiksi sanoja *sky*, *sunshine* ja *cloud* yhdistää yhteinen luontoon tai säähän liittyvä merkityskenttä. Kollokatiivinen koheesio perustuu heidän mukaansa sanojen merkitykseen eikä niinkään toistuvuuteen tai tilastolliseen merkitsevyyteen.

¹⁶ Käsillä olevassa tutkimuksessa *saneella* viitataan yksittäiseen sananmuotoesiintymään (vrt. engl. *token*) ja *sanalla* tai *lemmalla* eri sananmuotojen muodostamaan abstraktioon (vrt. engl. *type*).

simman monipuolisesti ottaen huomioon niin leksikaaliset kuin syntaktisetkin rakenteet. Kieliopilliset kategoriat esitetään kontekstuaalisia piirteitä tutkivissa analyyseissa *kolligaahteina*, kuten sanaluokkina tai lauseenjäsenluokkina.

Kollokaation tavoin myös kolligationaalisen suhteen on esitelty Firth, joka on määritellyt sen erilaisten *kielioppikategorioiden* keskinäiseksi suhteeksi:

The statement of meaning at the grammatical level is in terms of word and sentence classes or of similar categories and of the interrelation of those categories in colligations. Grammatical relations should not be regarded as relations between words as such – between ‘watched’ and ‘him’ in ‘I watched him’ – but between a personal pronoun, first person singular nominative, the past tense of a transitive verb and the third person pronoun singular in the oblique or objective form. (Firth 1957: 181.)

Firth tarkoittaa kolligatiolla siis sanaluokkien, lauseenjäsenluokkien tai erilaisten sananmuotojen yhteisesiintymistä syntagmaattisessa suhteessa.¹⁷ Kolligatiion käsite on tästä kuitenkin laajentunut tarkoittamaan myös muunlaisia kuin vain kielioppikategorioiden keskinäisiä syntagmaattisia suhteita. Kolligatiiona voidaan pitää ensinnäkin *yksittäisen lekseemin* ja jonkin kontekstissa olevaan kielioppikategorian suhdetta (mm. Hoey 1997: 8; Sinclair 1998: 15; Stubbs 2001b: 449). Tästä antaa esimerkin Sinclair, joka on havainnut yhdeksi BUDGE-lekseemin (‘hievahtaa’) merkittäväksi kolligatioksi kombinaation BUDGE <modaaliverbi> (mts. 13).¹⁸ Nykyisin – etenkin korpuslingvistiikassa – kolligatiolla tarkoitetaan useimmiten juuri lekseemin ja kielioppikategorian syntagmaattista suhdetta (Tognini-Bonelli 1996: 74).

Kolligationaalisia suhteita on tarkasteltu myös lekseemitasoa laajemmasta ja suppeammasta näkökulmasta, ja myös tällöin kuvauksissa on pitäyditty kolligatiokäsitteessä. Sinclair (1998: 15) pitää kolligatiiona myös lausekkeiden suhdetta kontekstin kielioppiluokkiin. Leksikaalisia yksiköitä analysoidessaan hän on käyttänyt esimerkkinä muun muassa lauseketta *naked eye* (‘paljas silmä’), joka saa vasemmanpuoleiseksi vieruskolligatiikseen yleisimmin määräisen artikkelin ja sitä edeltäväksi kolligatioksi preposition. Näin tämä leksikaalinen yksikkö koostuisi lausekkeen ja ainakin kahden kolligatiin syntagmasta: *naked eye* <prepositio, määräinen artikkeli>, esimerkiksi *just visible to the naked eye* (Sinclair 1996: 85). Toisaalta myös lekseemin eri sananmuotojen on havaittu muodostavan omia kolligationaalisia suhteitaan. Tognini-Bonellin (1996: 74, 77–80) mukaan esimerkiksi italian samaa tarkoittavilla ja näennäisesti vaihtoehtoisilla infinitiivimuodoilla *saper* ja *sapere* (‘tietää’) on omat syntagmaattiset valintansa: *saper* esiintyy infinitiivien mutta *sapere* sivulauseiden yhteydessä. Sinclair (1991: 8) ehdottaakin, että yksittäisiä sananmuotoja voisi pitää jopa omana leksikaalisena yksikkönään silloin, kun

¹⁷ Firthin määritelmään ovat nojautuneet myös Hakulinen ja Ojanen (1976: 81), joiden mukaan kolligatio on sanaluokkien tai sanaluokkien alakategorioiden muodostama suhde.

¹⁸ Toisinaan lekseemin ja kielioppikategorian suhdetta ei ole kuitenkaan pidetty *täytenä* kolligatiiona, jolloin tulkintaa ohjaa edellinen Firthin määritelmä (ks. esim. Sinclair 1996: 85).

sananmuodoille on osoitettavissa oma kontekstuaalinen valintansa. Myös Aston ja Burnard (1998: 8) pitävät mahdollisena sitä, että yksittäisellä sananmuodolla tai muotoryhmällä on oma kolligationaalinen valintansa; lisäksi he pohtivat myös mahdollisuutta, että lekseemin eri merkityksillä olisi omat kolligationaaliset esiintymäympäristönsä. Tämän on osoittanut todeksi muun muassa Hoey, joka on havainnut, että englannin REASON_{sy} saa kolligaateikseen useimmiten määräisen artikkelin (*the reason*), mutta kollokaatio REASON_{järki}, päättelykyky <*the*> on erittäin harvinaisen (1997: 9, 12). Mauranen (1998b, 2002a) on puolestaan todennut käännösaineiston perusteella, että yhden lemman eri sananmuodoilla on omanlaisensa merkitysprofiili. Tutkimuksessa paljastui esimerkiksi, että englannin THINK-lemman *think*-sananmuotoa käytetään yleisimmin epävarmuutta kuvaavassa merkityksessä 'uskoa, luulla', kun taas *thought*-sananmuotoa käytetään mentaalista prosessointia kuvaavissa merkityksissä, kuten 'ajatella'. Tällainen sananmuotojen merkitysvariantti lienee myös yksi peruste sananmuotojen erilaisille kollokationaalisille preferensseille.

Lekseemien eri merkitysten syntagmaattisten suhteiden tutkimisen lisäksi Hoey (1997: 8) korostaa myös sanojen lauseaseman tarkastelua, sillä hän pitää lauseasemaa yhtenä kolligaatioanalyysiin kuuluvana piirteenä kielioppikategorioiden tutkimisen rinnalla. Hoey havainnollistaa tätä käyttäen edelleen REASON-esimerkkiä: hän on havainnut aineistonsa pohjalta, että kombinaatio REASON_{järki} <persoonapronomini_{poss}> (esim. *my reason*) ei esiinny lauseen alussa mutta kylläkin muissa asemissa (mts. 13). Tapauksessa yhdistyy siten sekä itse tarkasteltavan sanan ja sen merkityksen että myös koko kolligaation suhde lauseasemaan. Myös Kennedy (1991) on tarkastellut lauseasemaa englannin *between*- ja *through*-lekseemejä analysoidessaan. Hänkin pitää lausepositiota yhtenä merkitykseen kuuluvana piirteenä ympäristössä olevien sanaluokkien rinnalla: esimerkiksi *through*-lekseemin oikeanpuoleisessa ympäristössä lauseenloppuisuus on merkittävin kolligaatti (mts. 102)¹⁹. Itse olen myös havainnut lauseposition olevan yksi kontekstuaaliseen merkitykseen liittyvä ominaisuus tarkastellessani lähisyronyymejä TÄRKEÄ ja KESKEINEN (Jan-tunen 2001b; tutkimuksesta tarkemmin luvussa 4.2).

Edellä käsitellyissä määritelmissä ja tutkimuksissa on kolligationaalisen suhteen pidetty tai sellaiseksi ehdotettu

- a) lekseemin
 - b) lekseemin eri sananmuotojen tai sananmuotoryhmien
 - c) useamman kuin yhden lekseemin muodostaman lausekkeen
 - d) yhden lekseemin eri merkitysten
- suhdetta johonkin kontekstissa esiintyvään kieliopilliseen kategoriaan tai vaihtoehtoisesti lausepositioon. Näistä on tähän asti tyypillisin tarkastelunäkökulma ollut lekseemien suhde kontekstin kategorioihin, mutta myös muilla lähestymistavoilla

¹⁹ Kennedyn (1991: 100, 102) tutkimuksessa ei varsinaisesti puhuta *kolligaateista* vaan kollokaatit on ryhmitelty sanaluokittain tai niitä pienempiin luokkiin. Käytännössä kyse on kuitenkin kolligaatioanalyysistä. Toisinaan kolligaatio-käsitteen rinnalla tai sijasta on käytetty myös sellaisia käsitteitä kuin *kieliopillinen kollokaatio* (mm. Carter 1987) tai *kieliopillinen assosiaatio* (Biber ym. 1998), joilla on tarkoitettu samaa ilmiötä.

voidaan saavuttaa olennaista tietoa lekseemien kolligationaalisista myötäesiintymisistä kuten jo aiemmin esitetyt tutkimukset osoittavat.

2.3.4. Kontekstin semanttinen valinta

1990-luvulla tutkimuskysymykseksi on noussut se, onko lekseemeillä (tai yksittäisillä sananmuodoilla) samankaltaista preferenssiä tietynlaisiin semanttisiin piirteisiin kuin niillä on kollokaatteihin ja kolligaatteihin. Näin myös kontekstissa olevien kollokaattien merkityssisältö ja tilannekontekstista nouseva merkitys on liitetty leksikaalisen yksikön valintapiirteiden kuvaukseen. Kyseessä on siirtyminen kollokaatiota abstraktimmalle tasolle: kun kontekstissa olevat kollokaatit voi konkreettisesti havaita ja poimia, niiden semanttiset ominaisuudet on tutkijan puolestaan pääteltävä ja luokiteltava erilaisten semanttisten kategorioiden avulla (vrt. myös kollokaattien ja kolligaattien suhteeseen). Samalla myös intuition määrä lisääntyy analyysissa: viime kädessä vasta aineiston analysoija pystyy määrittelemään kollokaattien merkityssisällöistä ilmenevät semanttiset ominaisuudet, joita korpuslingvistiikassa on nimetty semanttiseksi preferenssiksi ja semanttiseksi prosodiaksi.

2.3.4.1. Semanttinen preferenssi

Semanttinen preferenssi on edellä mainituista kahdesta kontekstuaalisesta merkitysominaisuuksista harvemmin käsitelty, ja joskus sellaiseksi luettavaa valintaa ei ole erotettu semanttisesta prosodiasta. Sen on määritellyt ehkä selkeimmin Stubbs, vaikka ilmiön onkin esitellyt ensimmäisen kerran Sinclair (1996, 1998):

Semantic preference is the relation, not between individual words, but between a lemma or word-form and a set of semantically related words (Stubbs 2001a: 65).

Semanttinen valinta syntyy siis sananmuotojen tai lekseemien ja niiden kollokaattien merkityssisältöjen suhteesta. Myös tämän ominaisuuden on oltava säännönmukaista myötäesiintymistä, jotta leksikaalisella yksiköllä voidaan katsoa olevan omanlaisensa semanttinen valintapreferenssinsä (Sinclair 1998: 16). Semanttiselle preferenssille on ominaista, että sitä ilmentävät sanat eivät ole sidoksissa tiettyyn positioon noodin kontekstissa eivätkä toisaalta myöskään sanaluokkiin tai niitä pienempiin luokkiin. Tämä tulee ilmi Sinclairin (1996: 83–89) *naked eye* -lauseketta ('paljas silmä') koskevasta analyysistä, jossa kyseisen ilmauksen semanttiseksi preferenssiksi paljastuu 'näkyvyys'. Tutkimus osoittaa, että preferenssi tulee ilmi sekä kontekstin verbeistä SEE ja VIEW ('nähdä, katsoa') että adjektiiveista VISIBLE ('näkyvä') ja INVISIBLE ('näkymätön'), eli preferenssi ei ole sidoksissa kollokaat-

tien sanaluokkaan. Toisenlaisen esimerkin antaa Stubbs, jonka mukaan COMMIT (mm. 'jättää haltuun, passittaa, tehdä, siirtää') saa lähes 30 prosentissa kaikista esiintymisistään kollokaateikseen lekseemejä, joiden merkitys kuvastaa rikollista tai muuten epätoivottavaa toimintaa (CRIME 'rikos', MURDER 'murha', SUICIDE 'itsemurha') (2001a: 64). Stubbsin antamasta esimerkistä selviää, että kyseisellä lekseemillä on myös muita semanttisia preferenssejä tai ainakin kollokaatteja, joiden merkitys ei mahdu tähän merkitysryhmään; lekseemeillä voikin olla useammanlaisia preferenssejä eri merkitysryhmiin, ja preferenssien määrä myös vaihdellee lekseemeittäin.

2.3.4.2. Semanttinen prosodia

Edellistä semanttista piirrettä useammin kirjallisuudessa käsitelty semanttinen prosodia²⁰ on preferenssiä abstraktimpi kontekstuaalinen ilmiö, eräänlainen merkitysaura (Louw 1993: 157), joka ilmausta ympäröi. Sinclair korostaa ensinnäkin, että semanttinen prosodia heijastaa kielenkäyttäjän asenteita ja että sillä on pragmaattinen funktio. Näin ollen semanttinen prosodia yhdistää kotekstin ja kielenulkoisen kontekstin ja on eräänlainen leksikaalisen yksikön ”uloin kerros” suhteessa kontekstiin. Toisin sanoen semanttisella prosodialla on kosketuspinta tilannekontekstiin. Sinclairin mielestä semanttisen prosodian tehtävä on ohjata ilmauksen tulkitusta, sillä se kertoo, mikä on ilmauksen funktio. Niinpä ilman semanttista prosodia ilmauksella olisi vain abstrakti merkitys ilman konkreettista käyttöyhteyttä ja tarkoitusta. Sinclair päätyykin tulkinnessaan siihen, että funktio, joka viestillä halutaan olevan, ohjaa semanttisen prosodian valintaa, joka edelleen vaikuttaa muiden leksikaalisen yksikön valintapiirteisiin ja että semanttinen prosodia on varsinainen syy siihen, miksi jokin ilmaus on ylipäättään valittu. (Sinclair 1996: 87–88, 1998: 20.) Myös Stubbs (2001a: 65, 2001b: 449) korostaa prosodian yhteyttä asenteiden ilmiötuomiseen, ja koska prosodia kertoo puhujan suhtautumisen johonkin asiaan, se myös samalla ilmaisee syyn, miksi ilmaus on tuotettu.

Aineistoon perustuen semanttista prosodiaa on käsitelty preferenssin tapaan ensimmäisen kerran juuri Sinclair (1991) tarkastellessaan verbirakenteen *set in* ('alkaa') kontekstuaalista valintaa. Aineistonsa perusteella Sinclair esittää, että verbirakenteen kontekstissa olevat subjektit viittaavat yleensä epämiellyttäviin asioihin, ku-

²⁰ Stubbs (2001a: 65–66) huomauttaa, että ilmiön kuvaamisessa on edelleen olemassa terminologista variaatiota. Samassa yhteydessä hän pyrkii selkeyttämään kahden semanttisen kontekstuaalisen ominaisuuden eroa ja päätyy ehdottamaan semanttisen prosodian sijalle käsitettä *diskurssiprosodia* (ks. myös Tognini-Bonelli 1996: 209). Käsitteen valinnalla Stubbs haluaa korostaa viestin tuottajan ja vastaanottajan vuorovaikutusta viestintätilanteessa ja prosodian roolia diskurssin koherenssin luojana. Näkemys ei ole kuitenkaan aivan uusi, sillä Sinclairin semanttiseen prosodiaankin sisältyy oletus prosodian syntymisestä juuri vuorovaikutuksessa. Semanttinen prosodia ilmentää nimenomaan tilannesidoksista funktionaalista valintaa, syytä, jonka vuoksi kielenkäyttäjä on valinnut ilmauksen tekstiyhteyteen eli diskurssiin.

ten mätänemiseen (ROT), pahanoloon (MALAISE), tunnottomuuteen (NUMBNESS) ja anarkiaan (ANARCHY), vain muutama esimerkki tässä mainittuna. Ilmiötä on kuvattu usein käsiteparilla negatiivinen–positiivinen: noodilla todetaan olevan joko negatiivinen tai positiivinen prosodia, mikä ilmentyy kollokaattien kielteisinä tai myönteisinä merkityssisältöinä. Esimerkiksi Stubbs (1995a: 42–48, 1995b: 247–250, 1996: 127, 172–181) luettelee englannin kielestä lukuisia lekseemejä, joilla on kielteinen semanttinen prosodia. Niistä mainittakoon lekseemit CAUSE (‘aiheuttaa’) ja AFFECT (‘vaikuttaa’) muutamane negatiivista prosodiaa edustavine kollokaatteineen:

CAUSE <DEATH, PROBLEM> (‘kuolema, ongelma’)
AFFECT <BADLY, DISEASE> (‘pahasti, sairaus’).

Sen sijaan lekseemeillä PROVIDE ‘hankkia, tarjota’ ja CAREER ‘ammattiura’ on Stubbsin mukaan selvä positiivinen prosodia:

PROVIDE <CARE, HELP> (‘huolenpito, apu’)
CAREER <BRILLIANT, PROMOTION> (‘loistava, ylennys’).

Prosodiaksi on luettu myös muita merkityspiirteitä. Louw (1993) puhuu negatiivisen prosodian rinnalla myös ‘huonouden’ ja Stubbs (1995b) ‘epämiellyttävyiden’ prosodiasta, jotka ovat toisilleen hyvin läheisiä, ja Sinclair (1996) vuorostaan ‘vaikeuden’ prosodiasta.

Stubbs (2001a: 106) on nostanut esiin metodologisia ongelmia, joita tulee vastaan lekseemien prosodisia valintoja tutkittaessa. Stubbs huomauttaa nimittäin, että prosodian hahmottamiseen tarvitaan usein suurempi tarkastelualue kuin se, mikä saattaa riittää kollokaattianalyyseissa. Näin ollen tarkastelualueita voidaan joutua laajentamaan esimerkiksi kollokaatioanalyyysiin sopivasta alueesta, sillä sanat, jotka ilmentävät kulloistakin prosodiaa, voivat löytyä ennalta määrättyä aluetta kauempaakin (tosin on muistettava, että tarkastelualue on aina kompromissi myös kollokaattianalyyseja tehtäessä, mikä tuli ilmi jo aiemmin). Toinen ongelma tulee esille Stubbsin (mp.) esimerkissä, jossa tarkastellaan LAVISH-lekseemin (‘tuhlailevainen, ylenpalttinen’) kollokaatteja ja niiden ilmentämää prosodiaa: kollokaatti *parties* (‘juhlat’) itsessään ei kerro, onko *lavish parties* (yksi Stubbsin esimerkeistä) ilmaus, jossa tarkoitetaan ‘juhlia, joissa rahaa syydetään turhuuksiin’, vai ‘juhlia, joissa vieraita kestitään anteliaasti’. Tämä selviää vain laajemmasta tekstiyhteydestä – ja korpusanalyyseissa tarpeeksi laajalta konkordanssiriviltä. Prosodia on siis hyvin kontekstisidonnainen ilmiö, ja tulkinta jää analyysoijan pääteltäväksi.

2.3.5. Muotovalintoja eli morfologinen preferenssi

Edellä kolligaation tarkastelussa kävi ilmi, että myös lekseemin yksittäisillä taivutusmuodoilla voi olla oma kontekstuaalinen valintansa. Jo Firth (1968 [1957]: 181) on pitänyt tärkeänä yksittäisten sananmuotojen tutkimista, ja myöhemmin sekä Sinclair (1991: 8) että Stubbs (2001a: 27) ehdottavat, että leksikaalinen yksikkö voi muodostua yksittäisestä sananmuodosta, jolla on omanlaisensa kontekstuaalinen valinta. Lisäksi Sinclair (mp.) painottaa, että sananmuotoja pitäisi käsitellä lemmana vain jos niiden kontekstuaaliset ominaisuudet ovat tarpeeksi yhteneviä. Kontekstuaalisten morfologisten piirteiden tutkimus on kuitenkin ollut edellisiin piirteisiin verrattuna vähäistä, vaikka eittämättä myös morfologisia valintapreferenssejä on osoitettavissa, kuten seuraavasta käy ilmi.

Kun tutkimuksessa aloitetaan sanahaku aineistosta, on tehtävä päätös, haetaanko aineistosta tietoa lemman eli kaikki taivutusmuodot sisältävän abstraktion (vrt. lekseemi) vai yksittäisen sananmuodon eli lemman konkreettisen esiintymän avulla.²¹ Vaikka varsinaisesti sananmuotoihin ja niiden kontekstuaalisiin ominaisuuksiin porautuvia tutkimuksia on suhteellisen vähän, sananmuotojen kontekstuaalisiin piirteisiin on kiinnitetty huomiota joko muun tutkimuksen ohessa tai niiden merkitystä ei ole erikseen tutkimuksessa korostettu vaan sananmuotojen merkitys tulee tutkimuksessa esiin implisiittisesti lukijan purkaessa analyysimenetelmiä. Sananmuotokohtaisesta preferenssianalyysistä käy esimerkiksi kolligaatti-käsitteen tarkastelun yhteydessä (luku 2.3.3) mainittu Tognini-Bonellin italian SAPERE-verbin ('tietää') infinitiivimuojen erilainen kolligaattivalinta (1996: 77–80). Myös Stubbs on tarkastellut sananmuotojen kollokointia edelleen teoksessaan *Words and Phrases* (2001a, ks. myös 1996: 172–173), jossa hän esittää 200 miljoonan sanan korpuksen pohjalta esimerkin SEEK-verbin ('etsiä, tavoitella') eri muotojen kollokoinnista. Hänen analyysistään erottuu selvästi, että joillakin muodoilla voi olla yhteisiä frekventtejä kollokaatteja, mutta joillakin taas ei ainuttakaan: esimerkiksi *seek-* ja *seeking-*muotoja yhdistää 10 taajaan esiintyvää kollokaattia (kuten *ADVICE*, *GOVERNMENT*, *POLITICAL*) mutta muodot *seeks* ja *seeking* taas eivät saa yhteisiä frekventtejä kollokaatteja.

Morfologisesti rikkaasta suomen kielestä ei ole vaikea löytää tukea edellisille havainnoille. Tein omasta tutkimusaineistostani hakuja yleisillä adjektiiveilla KAUNIS ja HYVÄ, joiden sijamuodoista tarkastelin translatiivia ja ablatiivia. Näiden kontekstissa olevat yleisimmät verbikollokaatit on koottu alle. Ensiksikin *kauniiksi* saa kollokaateikseen puhumista ja muuttumista ilmaisevia verbejä lähes saman verran, mutta *kauniilta* taas selvästi eniten joltakin näyttämistä ilmaisevia verbejä:

kauniiksi 18 puhumista ilmaisevat: <SANOA, TUNNUSTAA, KEHUA, LUONNEHTIA, VÄITTÄÄ 50 %>
muuttumista tai muuttamista ilmaisevat: <TEHDÄ, MUUNTAA, MUUTTUA, KEHITYÄ, REMONTOIDA, LAITTA, LAANTUA 44 %>

²¹ Käsitteistä tarkemmin mm. Sinclair (1991: 41–42, 173) ja Stubbs (2001a: 25–27).

kauniilta 31 näyttää jonkinlaiselta: <NÄYTTÄÄ, VAIKUTTA 81 %>
kuulostaa jonkinlaiselta: <KUULOSTAA 6 %>

Myös *hyväksi-* ja *hyvältä-* sananmuodot erottuvat verbikollokaateiltaan selvästi toisistaan:

hyväksi 235 <OLLA 26 %, TEHDÄ 17 %, KÄYTTÄÄ 9 %, MUUTTUA 4 %, HAVAITA, TYÖSKENNELLÄ 3 %>

hyvältä 204 <TUNTUA 42 %, NÄYTTÄÄ 20 %, KUULOSTAA 13 %, MAISTUA, TUOKSUA 9 %, VAIKUTTA 3 %>

Ablatiivi *hyvältä* kollokoi *kauniilta*-sananmuodon tavoin NÄYTTÄÄ- ja KUULOSTAA-verbien kanssa, sekä erityisesti TUNTUA-verbin kanssa. Translatiiville *hyväksi* nämä eivät ole sen sijaan ollenkaan tyypillisiä. Molemmat esimerkit osoittavatkin, että lekseemien eri sananmuotojen kollokationaaliset preferenssit ovat selvästi toisistaan poikkeavat.²² Sekä englantia että tässä esitetyt suomea koskevat tutkimukset osoittavat, että eri sananmuodoilla voi olla erilaisia ja selviä kontekstuaalisia preferenssejä, jotka konkretisoituvat muun muassa erilaisina kollokaatioina. Tässä käsitelty morfologiaan liittyvä ominaisuus korostaa nimenomaan sitä, että lemman eri *sananmuodoilla* on oma kontekstuaalinen valintansa.

Toisaalta huomio voidaan kiinnittää myös kollokaattien morfologisiin piirteisiin. Tästä antaa erinomaisen esimerkin Kennedy (1991: 99–100) esitellessään englannin *between-* ja *through-*sanojen kollokationaalisia preferenssejä: hän toteaa muun muassa, että *between-*sanaa edeltävät nominit ovat yleisimmin yksikössä ja verbit pääasiassa menneen ajan partisiippi- ja finiittimuoja. Stubbs (1995b: 254) on puolestaan havainnut, että englannin *somewhat* ('jonkin verran') esiintyy usein komparaation sisältävissä myötäesiintymissä, kuten *somewhat* <*different, greater, better, older*>. Omassa, jo aiemminkin mainitussa TÄRKEÄ- ja KESKEINEN-adjektiivieja koskevassa analyysissäni oli havaittavissa myös selvää synonyymit toisistaan erotettavaa preferenssiä. Adjektiivien vasemmanpuoleisten substantiivikollokaattien sijamuotojakauma osoittautui tutkimuksessa selvästi erilaiseksi kolmen sijamuodon eli genetiivin, partitiivin ja elatiivin osuuskien suhteen. (Jantunen 2001b: 183–185.)²³ Morfologiseen valintaan liittyen Stubbs (2001a: 27–28) korostaakin, että tärkeää tietoa kontekstuaalisista piirteistä voi jäädä huomaamatta, mikäli tarkastelu jätetään lemman tasolle.

Edellä on kuvattu sananmuotojen ja kontekstin välisiä morfologisia preferenssejä, mutta käsittelyyn voidaan ottaa vielä kolmaskin morfologinen ulottuvuus, nimittäin hakusanan sisäinen morfologinen variaatio. Tätä on tutkittu erityisesti suomesta, ja nimenomaan korpuspohjaisessa synonyymiatutkimuksessa. Ensinnäkin

²² Kyseinen ilmiö liittyy lauseopeissa käsiteltyyn rektioon, jolla tarkoitetaan sanan määritteiden tai täydennysten sijavalintaa (ks. Hakulinen ja Karlsson 1979; Vilkuna 1996).

²³ Tutkimuksen vähyyden vuoksi on vielä epäselvää, miten hienojakoiseksi morfologinen analyysi kannattaa viedä. Onko kontekstuaalista preferenssiä löydettävissä esimerkiksi yhden morfeemin eri allomorfien (esim. talossaan/-nsa, mestarien/-den/-tten, veneisiin/-hin) väliltä, on täysin avoin kysymys.

synonyymiset TÄRKEÄ ja KESKEINEN ovat osoittautuneet toisistaan erottuviksi sijaituvuuden suhteen: TÄRKEÄ-adjektiivia on tutkimusaineiston mukaan käytetty eniten partitiivimuotoisena mutta KESKEINEN-adjektiivia nominatiivissa. Lisäksi nämä kaksi sanaa erottuivat toisistaan myös vertailuasteiden suhteen, sillä KESKEINEN on komparoimattomana selvästi useammin kuin TÄRKEÄ. (Jantunen 2001b.) Lekseemien morfologiaa on tarkastellut synonyymian näkökulmasta myös Arppe, joka on päätenyt omaa tutkimustani tukeviin tuloksiin. Tutkimuksessa, joka käsittelee muun muassa VOIDA- ja PYSTYÄ-verbien sananmuotojen preferenssiä, Arppe (2001) havaitsee, että PYSTYÄ suosii mennyttä aikaa ilmaisevia muotoja kun taas VOIDA passiivimuotoja. Samantapaista lähisyronyymien erilaista preferenssiä tiettyihin muotoihin on Arppe havainnut myös toisaalla käsitellessään POHTIA- ja MIETTIA-verbejä (Arppe 2002). Näitä tutkimuksia kuvataan tarkemmin synonyymisten kontekstuaalisuutta käsittelevässä luvussa 4.2.

Edellä kuvattujen havaintojen perusteella morfologiaan liittyvät havainnot voidaan jakaa kolmeen ryhmään:

- a) lemmän eri sananmuotojen suhde kontekstiin
- b) lemmän suhde kontekstissa olevien kollokaattien eri sananmuotoihin
- c) hakusanan sisäiset morfologiset valinnat.

Näistä kaksi ensin mainittua liittyy noodin ja kontekstin väliseen suhteeseen, jälkimmäinen lekseemin sisäiseen paradigmaattiseen preferenssiin. Hakusanojen morfologiaan valintoihin on siis olemassa joitakin viittauksia tähänastisissa tutkimuksissa, joskin havainnot ovat vähäisiä muista kontekstuaalista piirteistä tehtyihin huomioihin verrattuna. Silti näyttää selvältä, että lekseemejä ja niiden kontekstuaalista valintaa tutkittaessa ei pidä tyytyä tarkastelemaan lekseemejä ja kollokaatteja vain lemmoina; yksityiskohtaisempaa tietoa saadaan analysoimalla myös yksittäisiä sananmuotoja. Koska aiemmassa tutkimuksessa on löydetty selkeitä todisteita lekseemien paradigmansisäisestä preferenssistä ja sellaista on helposti löydettävissä morfologisesti rikkaasta suomesta, on aiheellista nostaa *morfologinen preferenssi* kollokaatioiden, kolligaatioiden ja semanttisten valintojen rinnalle leksikaalisia yksikköjä tutkittaessa. Tarvittaessa morfologinen preferenssi voidaan jakaa edellä kuvattujen mukaisesti *kontekstuaaliseen* ja *paradigmaattiseen* preferenssiin.

2.4. Kohti integroitua kontekstuaalista merkitysyksikköä

Edellä on kuvattu useita tähänastisia havaintoja siitä, millaisia systemaattisia yhteyksiä leksikaalisella yksiköllä saattaa olla kontekstinsa kanssa. Tällaiset kiinteät myötäesiintymät muodostavat leksikaalis-kieliopillisista ja semanttisista piirteistä koostuvan suhdeverkoston, joka kertoo leksikaalisen yksikön ”käytösäännöstön”. Leksikaalista yksikköä ei voi pitää kontekstistaan irrallisena, vaan pikemminkin se ja konteksti muodostavat kiinteän kaksisuuntaisen vuorovaikutussuhteen, jossa konteksti ohjaa leksikaalista valintaa ja toisaalta valittu leksikaalinen yksikkö toimii

kontekstina toiselle yksikölle (Sinclair 1998: 8). Kaiken kaikkiaan tekstintuottamisessa on kyse leksikaalisten yksiköiden ko-selektiosta (*co-selection*) eli yhteisvalinnasta. Tämä tarkoittaa, että yksittäisen leksikaalisen yksikön sijaan kielenkäyttäjä valitsee vähintään kaksi yksikköä (Tognini-Bonelli 1996: 81), jotka muodostavat valintarajoitusten ja preferenssien ohjaaman syntagmaattisen kokonaisuuden. Ilmauksen ja viime kädessä tekstin merkitys rakentuu paradigmaattisen ja syntagmaattisen dimension yhteensovittamisesta (Sinclair 1998: 14), eikä merkitysyksikkönä voi pitää yhtä sanaa vaan se muodostuu laajemmasta, kontekstuaaliset ominaisuudet sisältävästä leksikaalisesta yksiköstä. Sinclair päätyykin esittämään, että

[i]nstead of expecting to understand a segment of text by accumulating the meanings of each successive meaningful unit, here is the reverse; where a number of units taken together create a meaning, and this meaning takes precedence over the "dictionary meanings" of whatever words are chosen (mts. 6).

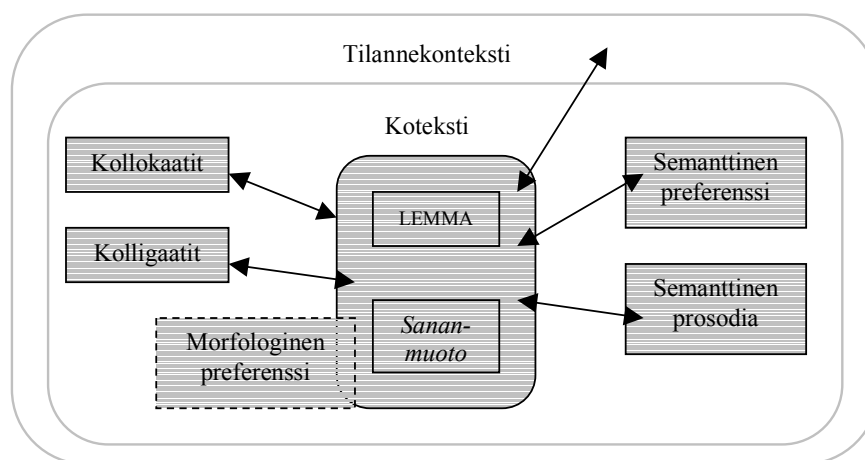
Viime kädessä konteksti siis ratkaisee leksikaalisen yksikön merkityksen tulkinnan eli luo sille tekstuaalisen merkityksen (Kenny 2001: 103). Merkityksen muodostuminen ei ole kuitenkaan tekstissä yksisuuntaista, sillä kun ympäristö vaikuttaa leksikaalisen yksikön merkityksen muodostumiseen, tuo sama yksikkö oman merkityspanoksensa tekstiin ja samalla vaikuttaa ympäristössään olevien sanojen merkitykseen.

Olisi kuitenkin virheellistä esittää, että teksti muodostuisi yksinomaan sanan ja kontekstin suhteesta. Jos palaamme jälleen aiempaan KIRJA- ja LUKEA-esimerkkiin (s. 18), jossa sanojen todettiin kollokoivan keskenään, voimme muodostaa esimerkiksi lauseesta *Luettuaan eilen illalla päivällisen jälkeen kirjaa Maija rupesi nukkumaan* monia muitakin kombinaatioita, joissa KIRJA ja LUKEA eittämättä kollokoivat mutta jotka eivät ole kieliopillisia: **Eilen illalla päivällisen jälkeen Maija nukkumaan luettuaan rupesi kirjaa*, **Maija kirjaa rupesi luettuaan eilen illalla päivällisen jälkeen nukkumaan*. Merkitysyksikön kontekstuaaliset ominaisuudet (yllä kollokaatio tai myös tilannekonteksti, jossa Maija lukee jotain kirjaa) eivät siis riitä yksistään muodostamaan sellaisia ilmauksia, jotka esiintyvät konventionaalissa kielenkäytössä. Valintarajoitusten ja -preferenssien lisäksi kielenkäyttötilanteessa vaikuttaa myös kielenpuhujien jakama tieto kieliopista ja sen mukaisista rakenteista. Sinclair (1991: 102–105) esittääkin, että leksikon ja syntaksin kuvausta ei voi täysin erottaa toisistaan vaan tekstin tuottaminen perustuu kompleksiseen kieliopin ja leksikon rajoitusten ja preferenssien noudattamiseen. Noudattamatta jättäminen tuottaa puolestaan epäidiomaattiseksi, epätavalliseksi tai luovaksi luettavaa kieltä.

Kontekstuaalisen merkitysteorian mukaan merkitysyksikkö ei koostu siis yksittäisestä sanasta tai edes sanan suhteesta muihin sanoihin, vaan leksikon, kieliopin, semantiikan ja kielenkäyttötilannetta ohjaavan tilannekontekstin muodostamasta konstruktiosta. Greenbaum (1970: 10–14) korostaakin, että integroidulla kontekstuaalisten piirteiden analyysillä (*integrated approach*) pystytään kuvaamaan merkitysyksikön kontekstuaalinen valinta perusteellisemmin. Olen pyrkinyt tiivistämään kontekstuaalisen merkitysyksikön muodostumiseen vaikuttavien perustekijöiden

suhteen edellä esitetyn perusteella kuvioon 2.2 (kuviossa eivät ole kuitenkaan mukana kieliopin aiheuttamat rajoitukset).

Edellä esitetystä on tullut ilmi, että tarkasteltavana olevan lekseemin eri sananmuodot voivat saada erilaisia kontekstuaalisia valintoja. Niinpä kuviossa on eritelty lemma ja sananmuoto omiksi yksiköikseen. Sekä kontekstuaaliset että kielenulkoiset kontekstuaaliset valinnat on esitetty kuviossa kaksisuuntaisena vaikutussuhteena, sillä siinä missä ko(n)tekstin on todettu vaikuttavan leksikaalisen yksikön valintaan ja merkitykseen, tuo myös kontekstiin upotettu leksikaalinen yksikkö oman merkityssisältönsä tekstiin ja ohjaa samalla muiden yksiköiden valintaa. Katkoviivoin esitetty morfologinen preferenssi poikkeaa muista siinä, että se koskee myös itse leksikaalista yksikköä (paradigmaattinen morfologinen preferenssi) sen lisäksi, että leksikaalisella yksiköllä on kontekstuaalinen morfologinen preferenssi. Se on siten nähtävä myös itse leksikaalista yksikköä koskevana, sisäisenä valintana.



Kuvio 2.2. Leksikaalisen yksikön kontekstuaalisen profiilin osatekijät.

Kielenulkoisen kontekstin muodostavat useat leksikaalisen yksikön valintaa ohjaavat tekijät. Kaiken kaikkiaan valintaa ohjaa koko Firthin kuvaama sosiaalinen tilannekonteksti, mutta sitä voidaan kuvata myös pienempinä valintaa ohjaavina tekijöinä. Partington (1998:17) esittääkin kollokaatioiden käsittelyn yhteydessä valintaa ohjaaviksi tekijöiksi muun muassa tekstilajin, rekisterin ja tyylin. Niinpä tekstilajit, joilla on erilainen kommunikatiivinen päämäärä (ks. Heikkinen ym. 2000: 66), kuten esimerkiksi romaani ja EU-säädös, ohjaavat erilaisiin valintoihin, jotka näkyvät kaksisuuntaisena valintasuhteena. Kontekstuaalisen valinnan ja rekisterin yhteyttä käsittelee jo 1966 John Sinclair, jonka mukaan leksikaalisten yksiköiden, kollokaatioiden ja klustereiden esiintymistä voivat säädellä myös tilannesidonnaiset kielenkäyttötavat (1996: 429), jotka myös heijastavat kielenkäyttäjien sosiaalista ja kulttuurikontekstia. Stubbs (1995b: 249) tähdentää vielä, että kollokaatiiden yleisyys voi vaihdella merkittävästi sen mukaan, mitä tekstit käsittelevät. Tämä tekstien aihepiiriä koskeva huomio liittyy läheisesti rekisterin ja tekstilajin

ohjaamaan valintaan, mutta ei ole kuitenkaan yhtenevä näiden kanssa. Edellisten lisäksi kollokationaalista valintaa ohjaavat vielä tyylilliset seikat: tyyppillisistä kollokaattipreferensseistä voidaan poiketa tyylillisten vaatimusten vuoksi esimerkiksi tavoiteltaessa innovatiivista, luovaa ja huomiota herättävää ilmausta (Haskel 1971; ks. myös Carter 1987 ja Kenny 2001). Vihdoin viimein kontekstuaaliseen valintaan voidaan lukea myös yksittäisen puhujan idiolektille tyyppilliset ilmaisutapaan liittyvät piirteet, jotka ohjaavat leksikaalisten yksiköiden käyttöä. Viimeksi mainitun ominaisuuden tarkastelu vie kuitenkin kauas korpuslingvistiikan perimmäisestä metodologisesta periaatteesta, jonka mukaan tarkoituksena on kuvata suuren tekstimassan avulla kielessä vallitsevia yleisiä lainalaisuuksia. Yksittäisen kielenkäyttäjän tyylin tarkastelu antaa kuitenkin mahdollisuuden kuvata variaatiota ja testata havaittuja lainalaisuuksia.

Käsillä olevan työn kontekstuaalisten merkityspiirteiden analyysissä on kontekstuaaliset piirteet rajattu seuraavasti: Kollokaatio on käsitetty toistuvaksi, ennalta määrätyltä tarkastelualueelta havaittavaksi kahden lekseemin myötäesiintymäksi. Kollokaatioanalyysi on tehty sekä tilastollisin menetelmin että esiintymistaajuuksien perusteella. (Analyysin tilastolliset menetelmät on kuvattu luvussa 5.4.2.) Kolligaationa pidetään puolestaan lekseemin ja sen kontekstista tavattavan kieliopillisen elementin suhdetta. Morfologisena preferenssinä tarkastellaan sekä paradigmaattista preferenssiä että lemmän ja kontekstin välistä kontekstuaalista preferenssiä. Semanttinen preferenssi ja semanttinen prosodia on käsitetty vuorostaan edellä kuvattujen aiempien tutkimusten tavoin.

3. KÄÄNNÖSSUOMI OMANLAISENA KIELIMUOTONAAN

Tässä luvussa käsittelen käännessuomen luonnetta ja merkitystä yhtenä suomen kielen muotona sekä korpuspohjaista käännesskielen tutkimusta. Luon aluksi katsauksen polysysteemiteoriaan ja sen avulla suomennoskielen merkitykseen suomalaisessa kulttuurissa. Tämän jälkeen käsittely siirtyy korpuspohjaiseen käännesskielen tutkimukseen, jota lähestytään niin sanottujen verrannollisten korpusten ja universaaliyhypoteesien ja erityisesti niiden problematiikan näkökulmasta.

3.1. Käännessuomi suomalaisen kulttuurin monijärjestelmässä

Käännessuomi on tuttu kielimuoto niin suomen kielen puhujille kuin suomen kielen tutkijoillekin. Tavallinen kielenkäyttäjätörmää suomennoksiin helposti jo heti aamusta esimerkiksi hammastahnatuubia puristaessaan, muropakettia käsitellessään ja aamiaisella sanomalehteä lukiessaan. Päivä päättyy usein vieraskielisten ja suomeksi tekstitettyjen tv-ohjelmien katsomiseen tai käännetyn kirjallisuuden lukemiseen. Ei liene liioiteltua sanoa, että suomennokset ja käännessuomi ympäröivät meidät koko valveillaoloajan. Myös tutkimuksen kohteena käännessuomi on ollut taajaan, erityisesti silloin kun on kosketeltu kirjakielen kehittymistä. Tämä on luonnollista, koska kirjakielen synty ja kehitys on ollut aina 1800-luvun loppupuolelle asti vahvasti käännessoiminnan varassa. Kirjakieltä tutkittaessa myös käännessnäkökulma on ollut tutkimuksessa läsnä: Esimerkiksi Agricolan kieltä eri aikoina tutkineet Ojansuu (1909) ja Itkonen-Kaila (1997) ovat ottaneet huomioon lähdekielen vaikutuksen syntaktisiin rakenteisiin, samoin sanaston ominaispiirteitä ja kehittymistä on tarkasteltu käännessvaikutusta silmällä pitäen (esim. Rapola 1960; L. Hakulinen 1969). Saukkosen (1984) tutkimus on puolestaan selittänyt tyylilajien syntyä käännesslähtöisesti. Laajamittaisesti kääntämisen vaikutusta kirjakielen kehitykseen on alettu tutkia myös Kotimaisten kielten tutkimuskeskuksessa meneillään olevassa 1800-luvun kirjakielen tutkimushankkeessa.

Suhteellisen uutta näkökulmaa sen sijaan edustaa viimeaikainen, laajaan korpusmateriaaliin perustuva vertaileva tutkimus, jossa kääntämisen tuloksena syntynyt suomennoskieltä verrataan niin kutsuttuun alkuperäissuomeen eli ”spontaanisti” suomeksi tuotettuun kieleen. Tässä tutkimuksessa näkökulma ei ole ollut kielen kehityksessä, vaan käännesskielen omanlaisessa erityislaadussa. Vertailevan tutkimuksen perustana on polysysteemi- eli monijärjestelmäteoria (Even-Zohar 1978, 1990 [1979]²⁴), jonka avulla käännessösten asemaa ja viime aikoina erityisesti käännesskieltä on tarkasteltu osana kohdekulttuurin ilmaisuvaroja. Teorian kulmakivenä on vaatimus siitä, että käännessöksiä, niiden kieltä tai jopa yksittäisiä ilmauksia ei

²⁴ Artikkelin ensimmäinen versio on painettu jo vuonna 1979 teoksessa *Poetics Today* 1(1).

pidä käsitellä irrallisina kokonaisuuksina, vaan ne on suhteutettava olemassa olevaan ympäröivään todellisuuteen. Even-Zoharin (1990 [1979]: 15) sanoin:

”Those interested -- in the "actual" constitution of products (e.g., lingual utterances, literary texts), cannot avoid taking into account the state of the particular polysystem with whose products they happen to deal.”

Omassa tutkimuksessani asetan käännösprosessin tuloksena syntyneen suomen kielen ilman kääntämistä syntyneen alkuperäissuomen rinnalle. Tällöin voidaan vertailla näitä kahta kielimuotoa ja analysoida käännössuomen erityispiirteitä.

3.1.1. Polysysteemi- eli monijärjestelmäteorian lähtökohdat

Itamar Even-Zoharin (1990 [1979]) *polysysteemiteoria*²⁵ on teoreettinen malli kulttuurin eri järjestelmistä, järjestelmien muutoksesta ja vaikutuksesta toisiinsa sekä käännöskirjallisuudesta osana kulttuurijärjestelmiä. Polysysteemi- eli *monijärjestelmäteorian* tavoitteena on kuvata kommunikaation eri toteutumia (kuten kulttuuria, kirjallisuutta ja kieltä) sisällöltään ja rajoiltaan muuttuvina ja toisiinsa kytkeytyneinä systeemeinä eli kuvattavan ilmiön hypoteettisina suhdeverkostoina (Even-Zohar 1990: 27) tarkkarajaisten, stabiilien ja itsenäisten ilmiöiden sijaan. Even-Zohar kuvaa polysysteemiteoriaa dynaamisen systeemien teoriaksi, joka ottaa huomioon muutoksen ja variaation mahdollisuuden. Järjestelmiä ei pidä kuvata itsenäisinä, historiallisesti irrallisina kokonaisuuksina vaan pikemminkin diakronisen kehityksen sisältävinä verkostoina. Niinpä kulttuurisia järjestelmiä kuvataan teoriassa sekä suhteessa toisiinsa että diakronisesti. (Even-Zohar 1990 [1979]: 10–12.)

Kulttuuri on nähtävä teorian mukaan eritasoisten monijärjestelmien verkostona, megapolysysteemisenä kokonaisuutena (Lambert 1997: 9), joka on jatkuvassa muutostilassa. Polysysteemit sisältävät itsessään myös polysysteemejä, esimerkiksi koko kulttuurinen monijärjestelmä sisältää vuorostaan kirjallisuuden monijärjestelmän, joka koostuu edelleen eri systeemeistä. Yhden polysysteemin sisältämien osasysteemien epätasa-arvo ja hierarkkisuus aikaansaa muutoksia diakronisella tasolla (Even-Zohar 1990 [1979]: 14), joten systeemien tila ei ole ajassa vakio vaan niiden asema voi muuttua monijärjestelmän keskustan (kaanonin) ja periferian (ei-kaanonin) välillä. Kulloinkin periferiaan luetut systeemit on luokiteltu usein kiinnostuksen ja tutkimuksen ulkopuolelle kuuluviksi, kun taas keskeiseksi katsotut ja prestiisiasemassa olevat ilmiöt – kuten kanonisoitu kirjallisuus ja kieli – ovat olleet tutkimuksen ja mielenkiinnon keskiössä. Polysysteeminen lähestymistapa kuitenkin

²⁵ Polysysteemiteoria pohjautuu pääasiallisesti venäläiseen formalismiin (mm. Jakobson ja Tynjanov), ks. Hermans (1999: 102–105). Lambertin (1997: 8) mukaan polysysteemiteoria on esitelty kirjallisuudessa ensimmäisen kerran Gideon Touryn julkaisussa *Li-she'elat te'ur ha-sifrut ke-rav ma'arekhet* (Literature as a Polysystem) teoksessa *Ha-sifrut/Literature* 18–19 (1974); pääsääntöisesti teorian luoja ja kehittäjänä pidetään kuitenkin Even-Zoharia.

painottaa tutkimuksen laajentamista myös ohi kanonisoitujen ilmiöiden, ja lähestymistapaa soveltavassa tutkimuksessa mielenkiinnon katsotaankin nousevan juuri keskustassa olevien ja perifeeristen ilmiöiden vuorovaikutuksesta, koska systeemin olemus selittyy sen asemasta verkostossa ja suhteista muihin systeemeihin. (Lambert mp.) Niinpä esimerkiksi kanonisoitua kirjallisuutta ei ole olemassa ilman ei-kanonisoitua eikä kanonisoitua kieltä ilman kanonisoimatonta. Se, mikä kussakin järjestelmässä, esimerkiksi kielen järjestelmässä, kuuluu keskiöön ja mikä periferiaan, ei määrity itsestään, vaan sijoittumista ohjaavat muut polysysteemin osat ja viime kädessä toiset polysysteemit, kuten yhteiskunnan poliittiset, uskonnolliset ja kulttuuriset monijärjestelmät.

Systeemit eivät ole suhteessa toisiinsa autonomisia eivätkä tarkkarajaisia. Sen sijaan ne vaikuttavat toistensa olemukseen ja toimintaan. Niinpä esimerkiksi kirjallisuuden monijärjestelmässä markkinat (yksi Even-Zoharin mainitsemista kirjallisuuden osajärjestelmistä) ohjaavat sitä, mitä julkaistaan ja mitä kulutetaan, toisaalta instituutiot taas sitä, mitä ylipäätään hyväksytään tuotettavaksi. Järjestelmien rajojen sumeus tulee erityisen hyvin esiin tarkasteltaessa kääntämisen ja käännöskielen systeemejä osana kirjallisuuden ja kielen monijärjestelmiä. Kuten Oittinen (1995: 53) sanoo, ”ennen kuin kääntäjä kirjoittaa ainuttakaan sanaa, hän on lukija”. Aloittaessaan kääntämisen kääntäjä (ja tulkkaja) siis vastaanottaa ensinnäkin yhteen kirjallisuussysteemiin kuuluvan teoksen eli kuuluu yhden kirjallisuussysteemin kuluttajasysteemiin lukijana. Kääntäessään tekstin kääntäjältä tulee toisen kirjallisuussysteemin jäsen, ja silloin hän on osa tuottajasysteemiä kirjoittajana. Käännösprosessin aikana kääntäjä on siis samanaikaisesti osa kahta eri (osa)systeemiä – ja kahden eri kirjallisen ja kielellisen monijärjestelmän jäsen. Aina ei myöskään ole helppoa erottaa käännöskieltä ja käännöstä alun perin jollekin kielelle kirjoitetusta tekstistä ja alkuperäiskielestä. Paloposki (2001) korostaa, että niin sanotut hybriditekstit ovat olleet tavallisia kirjallisuuden kehittyessä, mistä johtuu, että kahta eri tekstintuottamistapaa ja kahta eri kielimuotoa ei ole aina mahdollista erottaa toisistaan yhdessä teoksessa. Tämän voi katsoa pitävän paikkansa vielä nykyisinkin: mikä onkaan esimerkiksi tieteellisissä teksteissä kääntämisen tuloksena syntynyttä kieltä ja miten käännöskielen voi erottaa alkuperäiskielestä?

3.1.2. Käännössuomen ja suomennosten aseman muutos

Kääntämisen ja käännöskielen asema ei ole kulttuurissa stabiili, vaan se vaihtelee kulttuureittain ja aikakausittain. Even-Zohar (1978: 120–124) esittää, että käännökset ovat kirjallisessa monijärjestelmässä joko *primaarisessa* tai *sekundaarisessa* asemassa. Se, kumpaan asemaan kääntämisen systeemi asettuu, riippuu koko laajasta yhteiskunnan kulttuurisesta monijärjestelmästä. Primaarisen aseman käännöskirjallisuus saavuttaa silloin, kun a) kirjallisuuden polysysteemi on nuori ja vakiintumaton, b) kirjallisuus on perifeeristä tai heikkoa ja c) yhteiskunnassa on käännekohtia tai kriisejä tai siinä vallitsee kirjallinen tyhjiö (mts. 121). Ollessaan primaarisessa asemassa käännöskirjallisuus on kirjallisen polysysteemin keskiössä ja muok-

kaa vahvasti koko systeemiä. Silloin käännöskirjallisuus on laajasti innovatiivista suhteessa olemassa olevaan kirjallisuuskulttuuriin ja käännösten avulla esitellään kohdekulttuurille tuntemattomia ilmiöitä. Toissijaisessa asemassa ollessaan käännöskirjallisuus on puolestaan konservatiivista ja vahvistaa keskiössä olevan muun kirjallisuuden ominaispiirteitä.

Myös käännöskielen asemaa voidaan tarkastella edellä esitetyn jaottelun avulla. Hyvän esimerkin tästä tarjoaa käännössuomen asema suomen kirjakielen kehityksessä. Koska kirjoitettu suomen kieli syntyi sekä eli ja kehittyi pitkään suomennosten ja suomentamisen varassa, käännössuomi on ollut luonnollisesti hyvin primaarisessa asemassa kirjakielen kehityksessä. Käännetyt kirjalliset tuotokset olivat pitkään yhteiskunnallisesti merkittäviä lainsäätämiseen ja hengelliseen elämään kuuluvia tekstejä, joilla oli oma prestiisiasema muihin kirjoitettuihin teksteihin verrattuna. Tässä mielessä käännössuomi on siis ollut kirjoitetun kielen keskiössä ja kanonisoitua kieltä. Käännöskielen prestiisi näkyy selvästi raamatunsuomennosten kielen asemassa: ”Kun harvoja poikkeuksia lukuun ottamatta suomeksi liikuttiin vain kirkon ja uskonnon alueilla, tuli suomalaisesta piipiasta se koulu, josta sanonnan muodot, retoriset kuviot ja kuvakieli ratkaisevasti opittiin” (Rapola 1967: 105). Suomennokset täyttivät Suomessa siis yhtäältä suomenkielisen kirjallisuuden ja toisaalta suomen kirjakielen tyhjiön, ja niiden avulla, etenkin parannetuilla raamatunsuomennoksilla, pyrittiin kehittämään ja vakiinnuttamaan kirjakielen käytänteitä.

1800-luvulla syntynyt yhteiskunnallisen muutoksen mahdollistama kansallinen herääminen tarvitsi ravinnokseen myös muita kuin hengellisiä ja hallinnollisia tekstejä. Suomenkielisten sanomalehtien lisäksi alettiin kiinnittää huomiota kotimaisen kaunokirjallisuuden kehittämiseen, ja samalla sai käännöstoiminta osin uuden tehtävän: nyt suomennoksia tarvittiin kansanperinteen rinnalle malliksi kotimaiselle kaunokirjallisuudelle (Kovala 1999: 299). Vielä 1830-luvulle tultaessa suomen kieli todettiin kuitenkin niin köyhäksi ja vakiintumattomaksi, että korkeataimisen lyriikan ja taideproosan tuottaminen suomeksi koettiin vaikeaksi. Yhdeksi ratkaisuksi suomen kielen kehittämiseksi tarjottiin ulkomaisen kirjallisuuden, erityisesti klassikoiden suomentamista (Hellemann 1970: 419). Esimerkiksi Lönnrot ehdotti vuonna 1848, että *Kaunokirjallisuuden yhdyskunnan* olisi julkaistava kuu-kausittain yksi kaunokirjallinen suomennos (Kovala 1996: 42–43), ja häntä säesti samana vuonna myös Snellman *Litteraturbladissa*: ”kieltämättä onkin valittava tämä tie [käännöstoiminta], jotta voitaisiin luoda alkuperäiskirjallisuus, joka on nykyajan ja kansakunnan tietoisuudessa sivistyneemmän kirjallisuuden tasolla” (Hellemann 1970: 418–419). Suomennoksista tuli siten välineitä, joiden avulla pyrittiin kansallisen suomalaisen kulttuurin – kielen ja sitä kautta kirjallisuuden – kehittämiseen.²⁶ 1800-luku mukaan lukien siihenastinen aika oli ”käännöskautta” suomalaisen kirjallisuuden historiassa (Kovala 1999: 299), samoin luonnollisesti suomennoskieli

²⁶ Vaikka suomennosten avulla pyrittiin kehittämään suomen kieltä notkeammaksi, käännössuomea pidettiin vakiintumattomuutensa vuoksi usein myös heikkolaatuisena (ks. Hellemann 1970: 427–428), joten esimerkkiasema ei ole ollut ehdoton. Käännössuomen arvostukseen liittyvä tutkimus tarjoaisi mielenkiintoisia kysymyksiä; joitakin havaintoja nykypäivän käännössuomen arvostuksesta on löydettävissä tutkimuksista Vehmas-Lehto (1989), Tirkkonen-Condit (2002) ja Jantunen (2003, 2004).

piti asemansa kirjakielen kehittäjänä aina siihen asti kunnes kotimainen suomenkielinen julkaisutoiminta kehittyi.

Käännösuomen osuus kaikessa suomeksi julkaistussa kirjallisuudessa on vähentynyt selvästi kohti vuosituhannen vaihdetta, ja samalla suomennoskieltä on alettu käyttää myös yhä eriytyneemmissä tarkoituksissa, sillä käännöskirjallisuuden tehtävät muuttuivat ja eriytyivät lähestyttäessä 1900-lukua ja 1900-luvulla. Yhdeksi tehtäväksi vahvistui erilaisten lajityyppien kehittäminen: muun muassa viihdekirjallisuus ja toiminnallinen nuorisokirjallisuus esiteltiin nimenomaan käännösten kautta. Tärkeäksi suomentamismotivaatioksi tuli myös ajankohtaisuus: tavoitteeksi tuli suomentaa ajankohtaista ulkomaista kirjallisuutta ja mahdollistaa ulkomaisiin kulttuurivirtauksiin tutustuminen; lisäksi käännöstoiminta valjastettiin palvelemaan muun muassa työnväen-, raittius- ja naisasialiikkeiden tavoitteita. Kotimaisen suomenkielisen kirjallisuuden osuus tavoitti suomennosten osuuden 1910-luvun alussa, ja sen määrä kasvoi 20-luvun lopulle asti rinta rinnan käännöskirjallisuuden kanssa. Lähempänä nykypäivää 1980- ja 90-luvuilla suomennosten osuus kaikesta painetusta kirjallisuudesta on asettunut noin 10–18 prosenttiin ja suomenkielisestä painetusta kirjallisuudesta 13–23 prosenttiin. Sen sijaan kaunokirjallisuudessa käännösten nimiketuotannon osuus on ollut kahtena viime vuosikymmenenä jatkuvasti suurempi kuin kotimaisen kirjallisuuden: käännöskaunokirjallisuuden osuus on ollut noin 65 prosenttia, joskin vuosituhannen kahtena viimeisenä vuonna osuus on kasvanut peräti 72 prosenttiin. Sen sijaan niin sanottu ”käännöskartta” eli kulttuurit, joista käännetään, on yksipuolistunut jatkuvasti kohti 2000-lukua angloamerikkalaisen kulttuurin lisätessä jatkuvasti osuuttaan lähdekulttuurina ja englannin kielen lisääntyessä lähdekielenä. (Helleman 1970; Kuivasmäki 1996; Kovala 1999; Jalonen 1999; Minkkinen 2001.)²⁷

Kaiken kaikkiaan käännöskieli muodosti erittäin merkittävän osan kirjoitetusta kielestä pitkälti 1800-luvun puoliväliin ohi, jolloin omakielisen kirjallisuuden määrä alkoi kasvaa ja lopulta ohitti suomennosten määrän. Merkille pantavaa on kuitenkin se, että käännösuomi on yhä säilynyt primaarisessa asemassa monissa genreissä, ja nykypäivänä suomennosten merkityksen voi katsoa olevan edelleen merkittävä, vaikka käännöskirjallisuus ei yleisesti ottaen ole enää kirjallisuuden monijärjestelmässä aivan keskiössä. Tyypillistä on aseman eriytyminen eri aloilla. Kuten aiemmin tuli ilmi, kaunokirjallisuudessa suomennosten merkitys on edelleen erittäin suuri. Tämän lisäksi sellaiset tekstilajit kuten tietokoneohjelmien sisäiset opasteet ja tv-ohjelmien ja elokuvien ruututekstitykset ovat lähes yksinomaan suomennoksia, joten näissä tekstilajeissa käännöksillä ja käännösuomella on primaarinen asema. Sen sijaan sellaisia tekstityyppejä, joita ei tuotettaisi ollenkaan suomentamalla, on ehkä vähemmän tai ne ovat ehkä huomaamattomampia. Tällaisia voisivat olla Tirkkonen-Conditin²⁸ mukaan niin sanotut arkitekstit, esimerkiksi todistukset, ruokalistat, hartaustekstit ja kuolinilmoitukset, jotka syntyvät yleensä ilman kääntämistä.

²⁷ Käännöskirjallisuuden vaiheista ja merkityksestä Suomessa tarkemmin Helleman 1970; Kovala 1996, 1999; Chesterman 1998; Jalonen 1999; Lilius 2000; Paloposki 2000, 2002.

²⁸ Sonja Tirkkonen-Condit henkilökohtaisesti 2003.

3.1.3. Monijärjestelmäteorian kritiikkiä

Polysysteeminen lähestymistapa on saanut osakseen myös kritiikkiä. Etenkin käännöstutkimuksen piirissä on arvosteltu laajasti sitä, että näkökulma on siirtynyt liiaksi kohdesysteemiin, ja lähdetekstin ja -kulttuurin arvo on vuorostaan tutkimuskohteena vähentynyt. Muutosta on kuitenkin pidettävä välttämättömänä, kun tutkimuksen tavoitteeksi on tullut kuvata käännöstä suhteessa sitä ympäröivään kohdekulttuuriin. (Bassnett 1998: 128.) Vaikka kuvauskohteeksi nostetaankin käännös ja sen suhde kohdekulttuuriin, ei tutkimuksen tarvitse silti hylätä lähtökulttuurinäkökulmaa. Tästä käy esimerkiksi Sari Eskolan (2002) tutkimus, jossa hän vertailee käännössuomen ja alkuperäissuomen syntaktisia rakenteita. Tutkimuksessa on päähuomio näiden kahden kielimuodon – eli käännöksen ja kohdekulttuurin alkuperäiskielen – vertailussa mutta kielimuodon eroja selitetään nimenomaan lähdekielen eroilla suhteessa kohdekieleen ja toisiinsa. Lähdekieli (ja -kulttuuri) voidaan siis mainiosti yhdistää monijärjestelmäteoriastakin kumpuavaan tutkimukseen.

Kritiikkiä on saanut osakseen myös teorian perusteena olevien systeemien olemus. Lambert (1997: 11–12) ottaa esille kysymyksen siitä, ovatko systeemit todella olemassa olevia vai pelkkiä empiirisiä abstrakteja työkaluja. Nähdäkseni systeemit ovat sitä konkreettisempia ja havaittavampia, mitä pienemmistä ja tarkkarajaisemmista järjestelmistä on kyse, abstraktisuus taas lisääntynee suurempia kokonaisuuksia kuvattaessa. Esimerkiksi tuotteiden systeemi (kirjallisista tuotoksista muodostuva järjestelmä) on tarkkarajaisempi ja selvemmin havaittavissa oleva kuin vaikkapa teosten tuotantoa ohjaava markkinasysteemi. Kirjallisuuden systeemi on kokonaisuutena jo abstraktio, ja sitä suuremmat järjestelmät ovat yhä käsitteellisempiä. Käsitteellisyys ja järjestelmien avoimuus ei liene kuitenkaan tutkimuksessa ongelma, vaan hyväksyttävissä oleva tosiasia. Eivätkö mitkä tahansa kulttuuriset ilmiöt ole joka tapauksessa eräänlaisia abstraktioita, jotka manifestoituvat erilaisina ilmiöinä ja todentumina, esimerkiksi kieli eri muotoinaan, kirjallisuus tuotoksina ja niin edelleen?

Hermans (1999: 119) korostaa kuitenkin vielä, että systeemisen kuvauksen hierarkkisuuteen ja keskusta–periferia-suhteeseen perustuva jaottelu voi luoda illuusion siitä, että myös kuvauksen kohde olisi yhtä struktuurisesti järjestäytynyttä. Kääntämisen järjestelmän kuvaus lienee hyvä osoitus siitä, miten kuvattava kohde on ainakin jossain määrin sumearajainen ja järjestäytymätön. Määrittelyä vaatii aina esimerkiksi se, mitä pidetään käännöksenä (kaikki sellaiset tekstit, joita kulttuurin edustajat pitävät käännöksinä vai pelkästään tietyt kriteerit täyttävät tuotokset), kuka katsotaan kääntäjäksi (kaikki jotka jossakin määrin kääntävät, esim. kirjailijakääntäjät, vai koulutuksen saaneet ammatissa toimivat henkilöt) ja mistä roolista kääntäjän toimintaa tarkastellaan (vastaanottaja vs. tuottaja). Myös kieli on tarkastelun kohteena aina jossain määrin järjestymätöntä ja vailla selkeitä rajoja. Kun esimerkiksi verrataan alkuperäiskieltä ja käännöskieltä toisiinsa, on lähes mahdotonta tietää, missä määrin alkuperäiskieli on todella alkuperäistä ja originaalia. Koska tekstit ovat useimmiten jossain määrin intertekstuaalisia, sisältyy alkuperäiskieleenkin toisinaan sellaista materiaalia, jossa voi olla jonkin toisen tekstin vaikutusta ja jopa lainausta tai vieläpä jonkinasteinen käännöstoiminta taustalla. Tällai-

nen on tilanne usein lehtiteksteissä ja akateemisissa teksteissä, mutta mahdollisesti monessa muussakin kirjoittamisessa, jolloin kieli ei kaikilta osin saata ollakaan täysin ja tarkkarajaisesti alkuperäiskieleksi luettavaa. Tältä osin tutkimuksessa on tyydyttävä hyväksymään likimääräisyys ja kuvattavan kohteen rajojen sumeus.

3.2. Korpuspohjainen käännöskielen kuvaus

Käännöskieli on jätetty pois laajoista sähköisistä tutkimusmateriaaleista lähes systemaattisesti (Baker 1993: 234), mikä on jo riittävä peruste erityisten käännöskorpusten koostamiselle. Baker (mp.) olettaa, että käännökset on jätetty ulkopuolelle, koska niiden on katsottu vääristävän kuvaa tutkimuksen kohteena olevasta kielestä. Oletukselle löytyy tukea esimerkiksi Teubertin kategorisesta huomautuksesta:

Translations, however good and near-perfect they may be (but rarely are), cannot but give a distorted picture of the language they represent. Linguists should never rely on translations when they are describing a language (Teubert 1996: 247).

Käännöstekstien erottelu alkuperäiskielisistä teksteistä indikoi jo itsessään käännösten oletettua erityisluonnetta ja voisi sillään olla tutkimusmotiivi. Jos kielenkäytön kuvauksen pyrkimyksenä on kuitenkin tarkastella laaja-alaisesti kieltä, jonka kanssa kulloisenkin kielen käyttäjät ovat tekemisissä päivittäin, on lähes mahdotonta jättää käännetty kieli deskriptioon ulkopuolelle. Tämä koskee kaikkia niitä kulttuureja, joiden välillä kommunikointia harjoitetaan myös kääntämällä ja tulkaamalla. Näin ajatellen käännöskielen ja sen kuvauksen pitäisikin kuulua osana lukemattomien kielten deskriptioon ja sen vuoksi myös kielenkuvauksen perustana oleviin korpuksiin. Toisaalta kääntämisen tutkimuksella on myös erityistarpeensa, joihin yleiskielisillä korpuksilla, vaikka ne sisältäisivätkin käännöksiä, ei pystyttäisi vastaamaan. Yleiskieliset korpuksot antaisivat mahdollisuuden alkuperäiskielen ja käännöskielen vertailuun, mutta esimerkiksi lähdetekstien ja käännösten vertailu onnistuu vain tätä varten suunnitelluilla korpuksilla. Käännöskorpusten koonnilla on siis ollut ainakin kahdenlaiset motivaation lähteet: yhtäältä käännöksiä ei ole ollut tutkijoiden käytössä sähköisinä tekstimateriaaleina, ja toisaalta nimenomaan käännöstutkimuksen omalaatuisuus on luonut tarpeet materiaalin kokoamiselle.

Käännöskorpuksot ovat yleensä sisältäneet vähintään kahta eri kieltä (Baker 1995: 230); useimmiten tämä on tarkoittanut lähdekielisiä tekstejä ja niiden käännöksiä yhdelle tai useammalle kohdekielelle. Tällaiset niin sanotut *rinnakkaiskorpuksot* (*parallel corpora*) ovat hyödyllisiä sekä tutkimustyön aineistona että myös opetuskäytössä ja itse käännöstyössäkin, koska ne mahdollistavat aiempien, todellisten käännösratkaisujen tarkastelun. Myös *monikielisillä korpuksilla* (*multilingual corpora*) on sijansa kääntämisessä (erityisesti konekääntämisessä) ja opetuksessa: ne tekevät mahdolliseksi alkuperäiskielen tarkastelun todellisissa käyttötilanteissa, ei siis vain osana käännösprosessia. Kun kaksi edellistä korpustyyppiä sisältävät

vähintään kahta erikielistä aineistoa, niin sanotut *verrannolliset korpukset* (*comparable corpora*) ovat yksikielisiä. Baker määrittelee verrannollisen korpuksen rakenteeltaan sellaiseksi, joka sisältää kaksi osakorpusta samankielisiä tekstejä: osakorpuksista toinen sisältää käännöksiä jollekin kielelle ja toinen tekstejä, jotka on kirjoitettu suoraan samalle kielelle. (Baker 1995: 230–235.)

Vaikka sekä rinnakkaiskorpukset ja verrannolliset korpukset sisältävät molemmat käännöksiä, on niiden rakenne selvästi erilainen: verrannollisten korpusten sisältämät käännökset eivät ole käännöksiä toisen osakorpuksen teksteistä kuten rinnakkaiskorpuksissa, vaan nimenomaan käännöksiä yhdestä tai useammasta lähdekielestä sille kielelle, mitä toinen osakorpus edustaa. Verrannollisia korpuksia on luotu erityisesti alkuperäiskielen ja käännöskielen vertailuun. Näiden korpusten avulla voidaan löytää sellaisia piirteitä, jotka ovat erityisesti käännöskielelle ominaisia tai käännöksissä ali- tai yliedustettuja alkuperäiskieleen verrattuna. Bakerin luokittelun lisäksi käännöskorpuksia voidaan ryhmitellä edelleen sen mukaan, ovatko niiden sisältämät käännökset yhdestä vai useammasta lähdekielestä, julkaistuja vai julkaisemattomia, ammattilaisten vai esimerkiksi opiskelijoiden tekemiä tai onko käännökset tehty äidinkieleen vai vieraaseen kieleen päin. (Tarkempi kuvaus käännöskorpusten ryhmittelystä: Laviosa 1997, 2002; ks. myös Eskola 2002: 32–36.)

3.2.1. Käännössuomen korpus

Joensuun yliopiston kansainvälisen viestinnän laitoksessa koottu *Käännössuomen korpus* (Mauranen 1998a, 2000b) on edellä kuvatun kaltainen verrannollinen korpus, joka sisältää sekä alkuperäiskielisiä eli suoraan suomen kielelle tuotettuja tekstejä että näille verrannollisia käännösprosessin tuotoksena syntyneitä suomenoksia. Jotta korpusta voidaan pitää verrannollisena, on sen osakorpusten täytettävä kuitenkin muitakin kriteerejä kun vain se, että tekstit ovat samaa kieltä. Ensinnäkin aineiston tekstien on oltava osakorpuksittain samoista tekstilajeista. Käännössuomen korpuksessa (KSK) on edustettuna seitsemän tekstilajia: kaunokirjallisuus (kertomakirjallisuus), akateeminen kirjallisuus, lastenkirjallisuus, populaaritieteellinen kirjallisuus sekä muistelmala-, salapoliisi- ja viihdekirjallisuus. Näistä neljä ensimmäistä muodostavat laajimmat tekstilajikorpukset; kukin näistä neljästä sisältää yli miljoonan saneen tekstimassan. Toiseksi osakorpusten tekstit ovat peräisin samalta ajalta, muutamaa poikkeusta lukuun ottamatta 1990-luvun jälkipuoliskolta, joten aineisto on myös ajallisesti vertailukelpoista. Käännössuomen korpuksen kokonaissanemäärä on lähes 10 miljoonaa sanetta, noin 3,8 miljoonaa sanetta alkuperäissuomea ja noin 5,8 miljoonaa sanetta käännössuomea. Käännössuomen osakorpuksen lähdekieliä on kymmenen: englanti, espanja, hollanti, norja, ranska, ruotsi, saksa, unkari, venäjä ja viro, joista eniten on suomenoksia englannista ja venäjistä.

Korpuksen koostamiskriteereinä ovat olleet seuraavat seikat (Mauranen 1998a, 2000b, 2004a):

- a) tekstien autenttisuus eli tekstit ovat todellisia ja julkaistuja, eivät esimerkiksi kokeellisesti tuotettuja
- b) tekstien osittamattomuus, korpus sisältää kokonaisia tekstejä, ei fragmentteja
- c) professionaalisuus, käännökset ovat ammattikäytäntäjien tekemiä²⁹
- d) tekstin- ja kielenulkoiset kriteerit, eli yksittäisiä tekstejä ei ole valittu kielellisten piirteiden vaan ulkoisten luokittelujen (julkaisuluetteloiden) perusteella
- e) kulttuurinen merkitsevyys, mikä tarkoittaa, että tekstejä joko arvostetaan, luetaan tai käännetään paljon. Näin ollen tekstit ovat kulttuurisesti tyypillisiä, arvostettuja ja vaikuttavia.

Käännösuomen korpus ei ole laatuaan ainoa lingvistisesti orientoituneessa käännöstutkimuksessa käytetty verrannollinen tutkimusmateriaali. Alkuperäis- ja käännöskieltä sisältävää tutkimusaineistoa ovat hyödyntäneet mm. Puurtinen (1995) ja Vehmas-Lehto (1989) suomen sekä Gellerstam (1986) ruotsin ja soveltuvin osin Johansson (1998) norjan kielen osalta. Erona KSK:een on kuitenkin se, että tutkimusmateriaalien käännöskielen osuus on peräisin yhdestä lähdekielestä, Puurtisella, Gellerstamilla ja Johanssonilla englannista ja Vehmas-Lehdolla venäjältä – KSK:ssa lähdekieliä on taas useita. Ensimmäisenä varsinaisena elektronisena verrannollisena korpuksena on pidettävä Sara Laviosan koostamaa korpusta *English Comparable Corpus* (ECC; Laviosa-Braithwaite 1996, Laviosa 1998). ECC:n käännöskielen osakorpus perustuu UMISTissa Manchesterissa Isossa-Britanniassa koottuun korpukseseen *Translational English Corpus* (TEC), ja sen alkuperäiskielen osakorpus on peräisin British National Corpus -tekstipankista (BNC; Aston ja Burnard 1998). ECC perustuu siis aiemmin koottuun korpukseseen (BNC) ja sen rinnalle rakennettuun käännöksiä sisältävään korpukseseen (TEC). BNC ja TEC ovat koostamisperiaatteiltaan ja rakenteeltaan erilaisia, koska korpusten koostamisen taustalla olevat tavoitteet ovat olleet erilaisia. Tämän vuoksi näiden kahden korpuksen yhteismitallisuus on jossain määrin ongelmallista (ks. Laviosa-Braithwaite 1996: 75–81; Jantunen 2002: 112–114). Verrattavuuden ongelmat on pyritty KSK:n osalta eliminoimaan suunnittelemalla ja rakentamalla korpuksen molemmat osat samanaikaisesti ja samanlaisia koontiperiaatteita noudattaen. KSK onkin ensimmäinen varsinaisen käännöskielen ja alkuperäiskielen vertailua varten yhteismitalliseksi koostettu elektroninen tutkimusmateriaali.

3.2.2. Verrannollisten käännöskorpusten hyötyjä ja rajoituksia

Verrannollisten käännöskorpusten erityinen hyöty on luonnollisesti niiden mahdollistamana kääntämättömän ja käännetyn kielen vertailu, mitä varten verrannolliset korpuksot onkin ensikädessä luotu. Mutta verrannollisia korpuksia voidaan käyttää myös muihin tarkoituksiin. Lisäksi myös näillä, kuten kaikilla korpuksilla, on omat rajoituksensa.

²⁹ Joensuun yliopiston kansainvälisen viestinnän laitoksessa on koottu myös sivukorpusta, johon on kerätty opiskelijoiden harjoituskäännöksiä sekä alkuperäissuomalaisia tekstejä.

Ensiksi on kuitenkin syytä mainita, että verrannollisten korpusten koonti silläan osoittaa käännösprosessin tuloksena syntyneen kielen tutkimisen tärkeyttä kääntämättömän kielen rinnalla. Korpusten rakentaminen alkuperäiskielisten tekstikoelmien rinnalle implikoi, että käännetty kieli on jollakin tavalla muusta kielimateriaalista poikkeavaa ja että näiden erityispiirteiden tutkimus on kielenkuvauksen kannalta olennaista. Lingvistiikassa suhtautuminen käännöksiin on ollut paradoksisista: jättämällä käännökset pois yleiskielen korpuksista kielentutkijat ovat osoittaneet pitävänsä niitä jollakin tavalla muusta kielestä poikkeavana, mutta toisaalta samaan aikaan käännöksiä ei ole juurikaan pidetty tutkimisen arvoisina erilaisuudesta huolimatta. Tämä ajattelutapa on kuitenkin muuttumassa. Esimerkiksi Suomessa fennistit ovat yhä enemmän kiinnostuneet käännöksistä ja huomanneet niiden todellisen suuren merkityksen yhtäältä kirjoitetun suomen kielen kehityksen kannalta ja toisaalta käännösten suuren kulutuksen ja mahdollisen vaikuttavuuden vuoksi.

Yhtenä verrannollisten korpusten etuna voidaan pitää lisääntyvää objektiivisuutta tutkimuksessa. Niin kääntäjillä, kustantajilla kuin käännösten lukijoillakin on usein intuition perustuvia oletuksia siitä, millaisia käännökset ja käännöskieli ylipäätään ovat. Tällaiset subjektiiviset mielipiteet voivat perustua joko muutamien tekstien lukukokemukseen tai tutkimuksellisesta näkökulmasta muutamien lähde- ja kohdetekstien vertailuun, ja usein ne korostavat nimenomaan käännösten *toiseutta* alkuperäiskielen tai lähdetekstin rinnalla. Mauranen (2002b) tosin huomauttaa, että käännöstekstien arvostus saattaa määräytyä pikemminkin genren kuin alkuperän mukaan, sillä esimerkiksi klassikkoteoksia arvostetaan huolimatta siitä, että ne ovat käännöksiä. Pahimmillaan käännösten toisarvoinen asema näkyy käännösvirheiden kalasteluna. Verrannollisten korpusten olemassaolo ei tietystikään vähennä tällaisia vain muutamaa teksteihin perustuvia oletuksia, mutta ne kuitenkin mahdollistavat tutkimukset, jotka perustuvat laajoihin tekstimateriaaleihin. On kuitenkin hyvä muistaa, että objektiivisuus pätee tulosten tasolla, sen sijaan tulosten perusteella tehtävät päättelyt ovat aina subjektiivisia.

Objektiivisuus voi nousta ongelmaksi kuitenkin myös korpuksen koontivaiheessa. Tymoczkon (1998: 654) mukaan korpukset heijastavat tutkijoiden etukäteisodotuksia:

“corpora in translation studies are products of human minds, of actual human beings, and, thus, inevitably reflect the views, presuppositions, and limitations of those human beings”.

Tämä kysymys liittyy olennaisesti korpuksen edustavuuteen: millä kriteereillä korpuksen sisältämät teokset on valittu mukaan? Kysymys on tärkeä yhtä lailla käännöskielen tutkimuksessa kuin korpuslingvistikassa kielentutkimuksessa yleensäkin, sillä onhan valittujen tekstien tarkoitus edustaa tutkittavaa kielimateriaalia. Tärkeää on pohtia myös sitä, millä tavalla tuotettuja käännöksiä (kirjallisia käännöksiä vai tulkkausta, ihmisen vain koneen tekemiä käännöksiä jne.) ja mitä tekstilajeja tai -tyyppejä valitaan verrannolliseen käännöskorpukseen sekä millainen on käännösmateriaalin lähdekielten valikoima. Toisaalta on kuitenkin hyvä pitää mielessä, että

aineiston valintaan vaikuttavat aina tutkijan asettamat kysymykset ja tavoitteet, eikä korpustutkimus tästä eronne.

Verrannollisen korpuksen kokoamisessa voi tulla eteen Kennyn (1998: 53) kuvailema ongelma: kohdekulttuurin kirjallisuussysteemissä ei ehkä olekaan mitään käännöksille verrannollista materiaalia olemassa. Tällainen tilanne syntyy silloin kun käännösten kautta tuodaan jokin uusi tekstityyppi kohdekulttuurin systeemiin. Esimerkki on helppo löytää suomen kielen historiasta: suomen kirjakielen kehitys alkoi Mikael Agricolan käännöstyöstä, jolle ei ollut olemassa alkuperäissuomalaista verrannollista materiaalia. Tilanne jatkui verrannollisuuden kannalta problemaattisena tämän jälkeenkin: Paloposken (2002: 134) mukaan Agricolan jälkeen ilmestyneet teokset olivat eräänlaisia hybriditekstejä, jotka saattoivat sisältää sekä kääntämisen tuotoksena syntyneitä tekstimateriaalia että alun perin suomeksi kirjoitettua tekstiä. Pitkään käännökset ja alkuperäiskieliset tekstit eivät siis muodostaneet omia selkeästi erottuvia kategorioita vaan olivat läsnä yhdessäkin tekstissä, jolloin vertailtavien kategorioiden muodostaminen on vaikeaa.

Verrannollisia korpuksia on pidetty pelkästään teoreettisia tutkimuksia hyödyttävänä. Bowker (2000: 19) kritisoi, että niistä ei ole apua itse käännöstyössä. On kuitenkin muistettava, että verrannollisilla korpuksilla saatuja tutkimuksia voidaan hyödyntää kääntämisen opetuksessa ja tätä kautta tutkimuksella saavutettu tieto on myös kääntäjän löydettävissä (käännöskorpusten käytöstä kääntämisen opetuksessa ks. Kujamäki ja Jääskeläinen 2001, Kujamäki 2004). Lisäksi konkreettinen tutkimustieto on kääntäjän hyödynnettävissä erityisesti silloin kun hän haluaa – syystä tai toisesta – joko välttää käännöskielen ominaispiirteitä tai lisätä niitä tekstiinsä. Varsinaisessa käännöstyössä korpukset ovat lisäapu myös sanakirjojen ja kielioppien rinnalla, jotka yleensä tarjoavat tärkeää ”käyttöohjetietoa” sisältävää kontekstia hyvin vähän. Tirkkonen-Condit (2000a: 149–150) luettelee kolme käyttötapaa, joita korpusaineistoilla voi olla varsinaisen tutkimustyön lisäksi: a) kääntäjien apuvälinekäyttö; kääntäjä voi hyödyntää korpuksia ja niiden avulla tehtyjä tutkimuksia suomennosten tuottamisvaiheessa, b) leksikografinen käyttö; korpukset tarjoavat tekstuaalisesti sopivia vastineita ja kollokaatioita sanakirjan ilmoittaman tiedon rinnalle, ja c) pedagoginen käyttö; korpukset toimivat noviisikäntäjien apuvälineinä kääntämistä harjoiteltaessa. Lisäksi korpukset tarjoavat konkreettisia esimerkkejä todellisesta kielenkäytöstä, mistä voi olla apua erityisesti käännettäessä vieraaseen kieleen päin. Kaiken kaikkiaan monista soveltamismahdollisuuksista huolimatta on kuitenkin muistettava, että (verrannollisten) korpusten avulla tehdyn tutkimuksen ensisijaisena tarkoituksena ei ole kertoa kääntäjälle, millaista käännöskielen *pitäisi olla*, vaan kuvaus ja korpukset toimivat tietopankkina siitä, millaista käännöskieli *on*.

Viimeisenä otan esille lingvistiikan ja käännöstutkimuksen suhteen verrannollisten korpusten näkökulmasta. Fawcett (1998:120) väittää, että tieteidenvälinen tutkimus on hyödyntänyt selvästi käännöstutkimusta. Verrannollisilla korpuksilla tehtävä tutkimus todistaa kuitenkin myös päinvastaista: niillä tehtävä tutkimus ei tuota tietoa pelkästään käännöstutkimuksen tarpeiksi vaan kertoo samalla myös alkuperäiskielen ominaisuuksista. Tämä on väistämätöntä, koska vertaileva tutkimus tapahtuu kahden eri korpuksen – siis alkuperäiskielisen ja käännöskielen – välillä. Näin ollen alkuperäiskielisen korpuksen antama tieto on yhtä hyvin lingvis-

tien käytettävissä, ellei käännöskieltä koskeva tutkimus sillään tavoita mielenkiintoa. Tällaisia tutkimuksia on syntynyt jo englannista ja suomesta. Jälkimmäisestä mainittakoon Käännössuomi ja kääntämisen universaalit – tutkimus korpusaineistolla -projektissa tehdyt tutkimukset, jotka antavat käännöskieltä koskevaa informaation lisäksi relevanttia tutkimustietoa myös yleensä suomen kielestä: esimerkiksi Eskola (2002) on kuvannut suomen kielen syntetisoivia rakenteita, Mauranen (2000b) tieteellisten tekstien metatekstuaalisia piirteitä, Kemppanen (2002) historiatekstien sanastoa ja Nevalainen (2003) puhekielisten ilmausten käyttöä kaunokirjallisuudessa – vain muutamia projektin tähänastisista tuotoksista mainitakseni.

3.3. Oletuksia käännöskielen yleisistä tendensseistä

3.3.1. Käännöskielen universaalihypoteesit

Käännöskielen yleisiin tendensseihin kohdistuva tutkimus pohjaa yhtäältä kohdeteksti- ja kohdekielipainotteisen käännöstutkimuksen kehittymiseen ja toisaalta korpuslingvististen menetelmien rantautumiseen käännöstutkimukseen. Käännöstutkimuksen valtavirtaa on 1990-luvulle asti ollut lähtö- ja kohdetekstien ja lähtö- ja kohdekielen vertailuun sekä vastaavuuden ja ideaalikäännöksen ongelmaan keskittyvä tutkimus (Baker 1993; Tirkkonen-Condit 2000a). Kohdekielen ja erityisesti -kulttuurin konventioiden analyysi sai jalansijaa niin sanotun skopos-teorian myötä, joka tarkastelee kääntämisen funktiota ja adekvaattisuutta suhteessa kohdekulttuuriin ja käännöksen vastaanottajiin. Teoria muuttaa siis tarkastelukulman pois lähdetekstin ja käännöksen vertailusta käännöksen ja sitä ympäröivän systeemin tarkasteluun ja korostaa funktionaalista vastaavuutta. (Skopos-teoriasta Reiss ja Vermeer 1984; myös Vehmas-Lehto 1999: 90–98.) Niin sanotut deskriptivistit (Tirkkonen-Condit 2000a: 132–133) puolestaan tarkastelevat kääntämistä ja sen tuotoksia historiallisessa ja sosiokulttuurisessa kontekstissa tavoitteenaan kuvata kääntämisen tehtäviä ja normeja käännöksiä ympäröivässä kulttuurissa. Huomion kohteena ovat olleet niin poliittiset, esteettiset kuin taloudellisetkin intressit, jotka vaikuttavat kääntämiseen.

Korpuspohjaisen lähestymistavan kannalta näissä tutkimussuuntauksissa on olennaista empiirisen tutkimuksen kehittyminen ja arvostuksen lisääntyminen. Baker (1993: 237) korostaa myös käsitteellisen lähestymistavan ja lähde- ja kohdetekstien merkitysvastaavuuksien etsimisen väistymisen silottaneen tietä korpusstudiumille. Näiden tilalle on tullut ilmausten käytön kuvauksen näkökulma ja nimenomaan kohdekulttuurin käytänteiden huomioiminen. Muutos korostaa deskriptiivisen tutkimusotteen tärkeyttä ja on luonut tarpeen korpuspohjaiselle käännöstutkimukselle: korpuksen avulla pystytään kuvaamaan käännöksiä ympäröivää tekstuaalista (ja kulttuurista) todellisuutta ja siinä ilmeneviä käytänteitä. Niiden avulla voidaan vähentää myös käsitteelliselle ja lähdeteksti–kohdeteksti-vastaavuutta tavoittelevalle tutkimukselle tyypillistä introspektiota ja nojata todellista kulttuurista

tilannetta kuvaavaan aineistoon. Tällöin päästään nimenomaan Lambertin (1997: 9) vaatimien kulttuuristen lainalaisuuksien kuvaukseen. Erityisen soveliaaksi lähestymistavaksi Baker (mp.) hahmottelee käännostekstien ja niitä vastaavien kohdekielisten tekstien vertailun, mitä pystytään harjoittamaan nimenomaan verrannollisten korpuksen avulla.

Käännös-kohdekuulttuuri-vertailuun keskittyvä korpuspohjainen tutkimustraditio pohjaa aiemmin käsiteltyyn polysysteemiseen lähtökohtaan, jonka mukaan käännökset ovat osa laajaa kulttuurista monijärjestelmää. Even-Zohar (1990 [1979]: 13) korostaakin, että käännöksiä on tarkasteltava suhteessa muihin kohdesysteemin kirjallisuussysteemeihin eikä irrallaan niistä. Näin päästään kuvaamaan niitä ominaispiirteitä, jotka toisaalta erottavat kääntämisen systeemin muista systeemeistä ja toisaalta yhdistävät systeemit toisiinsa. Kohdekuulttuurin tarkastelua painottaa edelleen Gideon Toury (1985: 19): hänen mukaansa käännökset eivät ole enää osa *lähdekuulttuurin* systeemejä eivätkä myöskään vaikuta sen kielellisiin ja tekstuaalisiin käytänteisiin; sen sijaan ne tulevat käännösprosessin myötä osaksi *kohdekuulttuurin* systeemejä, jonka käytänteisiin, kuten kieleen, ne voivat myös vaikuttaa. Sovelletuna suomennosten tutkimukseen edellä kuvattu tarkoittaa sitä, että suomennoksia on tarkasteltava – ei pelkästään suhteessa lähtötekstiin vaan myös – suhteessa muihin suomalaisiin kirjallisiin ja kielellisiin systeemeihin, joihin käännösten voi olettaa vaikuttavan.

Verrannollisia korpuksia on tähän asti käytetty enimmäkseen niin sanottujen *käännösuniversaalien* tutkimuksessa. Käännösuniversaalit ovat oletuksia piirteistä, jotka ovat tyypillisiä käännöksille mutta eivät alkuperäiskielelle ja jotka eivät ole syntyneet minkään yksittäisen kielen vaikutuksen vuoksi. Universaaleiksi oletettujen piirteiden otaksutaan syntyvän väistämättä itse käännösprosessissa, ja niiden oletetaan manifestoituvan käännöskielessä erilaisina lingvistisinä piirteinä. (Baker 1993: 243; Shuttleworth ja Cowie 1997: 193–194; ks. myös Laviosa-Braithwaite 1998: 288–291.) Universaalipiirteillä voidaan siis tarkoittaa lainalaisuuksia (Eskola 2002: 45), jotka ilmenevät käännösten kielessä ajasta ja kieliparista riippumatta. Kääntämisen universaalien piirteiden olemassaoloa on ehdottanut Mona Baker (1993, 1996), joka perustaa oletuksensa aiemmin tehtyihin tutkimuksiin ja havaintoihin käännösten luonteesta.³⁰ Bakerin (1993, 1996) ja Laviosan (Laviosa-Braithwaite 1996, 1998; Laviosa 2002) esityksissä universaalihypoteeseiksi ovat hioutuneet *eksplisiittistyminen*, *konventionaalistuminen*, *tasoittuminen*, *yksinkertaistuminen* ja *epätyypilliset frekvenssit*, jotka voivat koskea leksikkaa sekä syntaktisia ja tekstuaalisia elementtejä. Tutkimuksen myötä hypoteesien kirjo on lisääntynyt: myös *epätyypillisiä leksikaalisia kombinaatioita* (Mauranen 2000b), *kohdekielisepesifisten uniikkiainesten aliedustumista* (Tirkkonen-Condit 2002, 2004) sekä *lähdekielen vaikutusta* (Mauranen 2000b, 2004a) on esitetty käännöskieliä yhdistäviksi

³⁰ Bakerin universaalihypoteesit kumpuavat yhtäältä Even-Zoharin (1978, 1990 [1979]) systeemisestä viitekehuksesta ja toisaalta Touryn (1991, 1995) esittämistä kääntämisen lainalaisuuksista. Lisäksi niiden taustalla on monta yksittäistä lähdeteksti-käännös-vertailulla tehtyä tutkimusta, tarkemmin Baker 1993, 1996. Universaaleiksi ehdotettujen piirteiden nimeämisperusteet ovat nähtävissä mm. Vanderauweran (1985) tutkimuksessa, jossa tämä käsittelee käännöksissä havaittavaa yksinkertaistumista, eksplisiittistymistä ja normalisoitumista (mm. mts. 73, 76, 97, 99).

piirteiksi. Universaalihypoteeseja on tähän mennessä esitelty kirjallisuudessa kattavasti (mm. Baker 1996; Laviosa-Braithwaite 1998; Laviosa 2002; Eskola 2002; Mauranten ja Kujamäki 2004), joten keskityn seuraavassa esityksessäni vain työtäni koskeviin hypoteeseihin. Näitä ovat oletukset käännöskielen epätyypillisistä frekvensseistä ja epätyypillisistä leksikaalisista kombinaatioista, yksinkertaistumisesta ja lähdekielen vaikutuksesta.

Epätyypilliset frekvenssit. Oletus käännöskielen epätyypillisistä frekvensseistä koskee niin käännösten leksikaalisia, syntaktisia kuin tekstuaalisiakin elementtejä; havaintoja on kuitenkin etupäässä kahdesta ensimmäisestä. Baker (1993: 245) ehdottaa, että epätyypillisten frekvenssien taustalla on itse käännösprosessi, joka muokkaa käännöskielestä omanlaisensa – sellaisen, jossa frekvenssit poikkeavat niin lähdekielen kuin kohdekielenkin vastaavista. Tässä yhteydessä Baker viittaa Frawleyn (1984) postuloimaan *kolmanteen koodiin (third code)*, millä tarkoitetaan nimenomaan käännösten kieltä omanlaisena kielimuotonaan, joka on syntynyt sekä lähde- että kohdekielen vaikutuksesta ja heijastaa molempien piirteitä, mutta poikkeaa kuitenkin näistä kummastakin. Toury (1980: 71–72) puolestaan viittaa käännöskielen kahden koodin välissä syntyvään omalaatuisuuteen käsitteellä *välikieli (interlanguage)*, jonka hän on lainannut vieraan kielen oppimisen teoriataustasta.

Esimerkkejä käännöskielelle ominaisten kielenainesten frekvensseistä ei ole vaikea löytää, sillä tältä osin tutkimusta on runsaasti. Tähän vaikuttaa se tosiasia, että monet universaalioletukset perustuvat ainakin joissain määrin epätyypilliseen kvantiteettiin, joko ilmiön ali- tai yliedustumiseen. Niinpä esimerkiksi englannin *that*-konjunktion yliedustuminen *SAY*- ja *TELL*-verbien objektiivisivulauseissa käännösendlannissa on tulkittu merkiksi eksplisiittisestä eli selkiinnyttävästä kielenkäytöstä (Olohan ja Baker 2000). Vastaavasti taas *jaksaa*, *mahtua*, *ehtiä*, *riittää* ja *viitsiä* -verbien, joilla ei ole leksikaalistunutta vastinetta esimerkiksi englannissa, aliedustuminen käännöksissä tukee uniikkien, vastineettomien aineiden vähyydestä esitettyä hypoteesia (Tirkkonen-Condit 2002, 2004). Syntaktisia rakenteita tarkastellut Eskola (2002) puolestaan pitää suomen kielen syntetisoivien rakenteiden (esim. lauseenvastikkeiden) alkuperäis-suomesta poikkeavia frekvenssejä osoituksena lähdekielen vaikutuksesta.

Edellä mainituissa viimeaikaisissa tutkimuksissa epätyypillisiä frekvenssejä ei ole pidetty epätoivottavina ilmiöinä käännöskielessä, vaan ne on tulkittu piirteiksi, jotka heijastavat käännöskielen inherenttejä ominaisuuksia. Käsitteet *tyypillinen* ja *epätyypillinen* onkin ymmärrettävä neutraaleina ja arvottomina ilmauksina, joiden avulla kuvataan käännöskielen olemusta, mutta ei aseteta käännöskieltä ja sille verrannollista kielimuotoa toistaan parempaan asemaan. Alkuperäiskielestä poikkeavia kieliainesten frekvenssejä on tarkasteltu aiemmin lähes yksinomaan arvottavassa kontekstissa. Esimerkiksi Vehmas-Lehdon (1989) tutkimuksessa syntaktisten rakenteiden poikkeavat frekvenssit todetaan *piilovirheiksi*, joiden välttämisen on suotavaa. Karhu (1997) puolestaan nimittää samaa ilmiötä *piilopoikkeamaksi*, joka määritellään kohdekielen käyttötavoista eroavaksi kvantitatiiviseksi poikkeamaksi, joka ei riko kohdekielen systeemisiä ominaisuuksia (mts. 72). Karhun käyttämä käsite on hyvin lähellä epätyypillisyyttä, koska ilmaus ei sisällä eksplisiittisesti virheen ilmipanoa kuten *piilovirhe*. Koska *piilopoikkeama*akin on kui-

tenkin pidetty vältettävänä käännöskielen ominaispiirteenä (mts. 85–86), on tyypillisyyss-käsitettä pidetty tässä suhteessa parempana käytettäväksi tutkimuksissa, joissa preskriptiivisen lähestymistavan sijalla on deskriptiivinen. (Ks. Jantunen ja Eskola 2002: 185; Eskola 2002: 59.)

Yksinkertaistuminen. Käännöskielen yksinkertaistuminen lienee tutkituin tähän asti esitetyistä hypoteeseista, sillä sitä on tarkasteltu sekä pelkästään yksinkertaistumiseen keskittyvissä tutkimuksissa (mm. Laviosa-Braithwaite 1996, 1997; Jantunen 2001a) että myös tutkimuksissa, joissa se on ollut yhtenä testattavana hypoteesina muiden joukossa (mm. Eskola 2002; Jantunen ja Eskola 2002). Oletus käännöskielen yksinkertaistumisesta perustuu lähdetekstien ja käännösten vertailussa tehtyihin havaintoihin. Näitä ovat muun muassa Blum-Kulkan ja Levenstonin (1983; ks. myös Laviosa 2002: 43–47) havainnot, joiden mukaan käännöksissä olisi alkuperäistekstejä yksinkertaisempi lekseemivalikoima. Tällaiseen tulokseen he ovat päätyneet tarkastelemalla hepreasta englantiin teytyjä käännöksiä; yksinkertaistumisen ilmentyminä he pitivät muun muassa yläkäsitteiden käyttämistä hyponyymien sijaan, tuttujen synonyymien suosimista ja parafraseja. Myös Vanderauwera (1985) on tehnyt sellaisia havaintoja, joita on pidetty yksinkertaistumista osoittavana: tutkiessaan hollannista englantiin käännettyjä tekstejä hän havaitsi, että alkuperäistekstien monimutkaisia lauserakenteita on pyritty käännöksissä yksinkertaistamaan esimerkiksi purkamalla infiniittirakenteita finiiittirakenteiksi (mts. 46).

Edellisten lisäksi Malmkjærin (1997) ja Mayn (1997) tutkimuksissa on esitetty havaintoja, joita voidaan pitää esimerkkeinä kielen ja tekstin rakenteen yksinkertaistamisesta. Tarkastellessaan H. C. Andersenin tuotannon englanninnoksia Malmkjær (mts. 152, 160) havaitsi käännöksissä välimerkitystä muutetun siten, että idiosynkraattinen tapa käyttää välimerkkejä oli muutettu englannin kielen konventioiden mukaiseksi ja yleensä siten, että merkkejä oli muutettu ”voimakkaammiksi”: pilkkuja puolipisteiksi ja pisteiksi ja puolipisteitä vuorostaan pisteiksi. Tällaista rakenteellista muokkaamista Baker (1996: 182) pitää strategiana, jolla tekstiä on pyritty muokkaamaan selkeämmäksi ja yksinkertaisemmaksi. Samankaltaisia havaintoja on tehnyt myös May, joka tarkasteli Virginia Woolfin ja William Faulknerin tuotantojen ranskannoksia ja venäjännöksiä: niissäkin välimerkitystä oli muutettu siten, että tekstistä tulee Mayn (1997: 7, 10) mukaan lukijalle helpompaa. Molemmissa tapauksissa alkuperäistekstien kirjoittajien epätavallista ja kokeilevaa kielenkäyttöä osoittavaa välimerkitystä oli muutettu yksinkertaisemmaksi ja norminmukaisemmaksi.³¹ Lisäksi Laviosa raportoi Shlesingerin (1991) havainnoista, joiden mukaan tulkki on vähentänyt ja poistanut tulkattavan puheessa esiintyneitä toistoja oikeussalitulkkauksessa. Tämän Laviosa tulkitsee tyylilliseksi yksinkertaistamiseksi. (Laviosa-Braithwaite 1996: 7; Laviosa 2002: 50.)

Universaalihypoteesi kielellisestä ja tekstuaalisesta yksinkertaistumisesta on siis luotu edellisten kaltaisten yksittäisten havaintojen perusteella. Varsinainen ensimmäinen laajaan korpusaineistoon perustuva hypoteesin testaus on Laviosan (Laviosa-Braithwaite 1996) väitöskirja, jossa hän tarkasteli ECC:n aineistosta

³¹ May (1997: 1) tulkitsee lähdetekstin ja käännöksen välimerkityksen erot osoitukseksi kääntäjien omaksumasta tekstin toimittamiseen pyrkivästä roolista.

koottujen sanomalehti- ja kertomakirjallisuuden korpusten avulla alkuperäisenglannin ja käännösenglannin eroja. Muuttujina Laviosa käytti lausepituutta, type-token-suhdetta (TT-suhdetta eli sananmuoto-sanesuhdetta) ja leksikaalista tiheyttä (leksikaalisten sanojen suhdetta kieliopillisiin sanoihin), joista ensimmäinen mittaa tyylillistä³² ja kaksi viimeistä sanastollista yksinkertaistumista. Lisäksi hän laski aineistosta mm. englannin 200 yleisimmän sanan peiton, millä myös pyrittiin tarkastelemaan leksikon rikkautta. Laviosan tulokset antavat jonkinlaista tukea yksinkertaistumishypoteesille: käännöksissä suosittiin englannin kielen tavallisimpia sanoja ja leksikaalinen tiheys osoittautui alhaisemmaksi käännöksissä. Sen sijaan koko aineiston sanastollista variaatiota mittaava TT-suhde ei antanut tukea yksinkertaistumishypoteesille. Samoin kävi lausepituuden suhteen, sillä lausepituus oli lyhyempi vain sanomalehtiteksteissä, ja kertomakirjallisuudessa lauseet olivat vuorostaan pitempiä. Tulokset antoivat siten vain jonkinasteista tukea yksinkertaistumishypoteesille.

Oma tutkimukseni (Jantunen 2001a), jossa tarkastelin käännösten leksikaalista yksinkertaistumista synonyymisten astemääritteiden avulla, ei antanut ollenkaan tukea yksinkertaistumishypoteesille. Tutkimieni astemääritteiden käyttö osoittautui yhtä varioivaksi tai jopa varioivammaksi käännössuomessa alkuperäissuomeen verrattuna. Eskola (2002: 264–265) sai pääasiassa syntaktisia rakenteita tarkastelevassa tutkimuksessaan osittain tukea yksinkertaistumishypoteesille: hänen mukaansa suomen kielen syntetisoivat ilmaisuvarat eivät ole käännöksissä yhtä laajasti käytössä kuin alkuperäiskielessä. Toisaalta hypoteesi ei kuitenkaan saanut tukea rakenteissa ilmenevää leksikaalista variaatiota tarkasteltaessa. Tulokset ohjaavat siis jälleen eri suuntiin. Kun lopuksi otetaan esille *hyvin*-astemääritteen kotekstin sanastollista rikkautta tarkasteleva tutkimus, jossa huomattiin kontekstuaalista leksikaalista yksinkertaistumista (Jantunen ja Eskola 2002: 201), havaitaan, että yksinkertaistumishypoteesia koskevat tulokset ovat hyvin ristiriitaisia. Osa tutkimuksista tukee hypoteesia, osa ei, ja jopa samassa tutkimuksessa on saatu vastakkaisia tuloksia. Eskola (mp.) huomauttaakin, että yksinkertaistumishypoteesi vaatisi entistä tarkempaa määrittelyä ja lisää konkreettisia tutkimustuloksia.

Epätyypilliset leksikaaliset kombinaatiot. Seuraavaksi käsiteltävä, Maurasen (2000b: 136) esittämä oletus käännöskielen epätyypillisistä leksikaalisista kombinaatioista perustuu Käännössuomen korpuksesta peräisin olevaan, akateemisilla ja populaaritieteellisillä teksteillä tehtyyn tutkimukseen, jossa hän vertasi sekä *toisaalta*-konnektorin että metatekstuaalisten verbien HALUTA ja PYRKIÄ kollokointia alkuperäis- ja käännössuomessa. Universaalihypoteesin postulaation perustana olivat muun muassa seuraavan kaltaiset havainnot: Alkuperäissuomessa tyypillisin *toisaalta*-konnektorin kollokaatio oli <mutta toisaalta>, mutta käännössuomessa <toisaalta taas> sekä <ja toisaalta>. Edelleen alkuperäissuomessa tavallinen kombinaatio HALUTA <korostaa> osoittautui harvinaiseksi suomennoksissa, jossa vuorostaan esiintyivät myötäesiintymät HALUTA <painottaa, osoittaa>; ne osoittautuivat vieraiksi vuorostaan alkuperäissuomelle. Alkuperäissuomelle epätyypillisten kombinaatioiden Mauranen (2000a) katsoo osaltaan rikastuttavan kieltä, mikä

³² Myös Saukkonen [1984: 61] on käyttänyt lausepituutta yhtenä tyylipiirteinä.

osoittaa, että niitä ei tarvitse luokitella vältettäväksi kielenaineiksi (vrt. edellä keskustelu *tyypillisuus*-käsitteestä). Maurasen esittämä hypoteesi on vastakohta niin sanotulle konventionaalistumisoletukselle (~ konservatiivistuminen ~ normaalistuminen), jolla on tarkoitettu kohdekielen ominaisuuksien yliedustumista ja tyypillisten piirteiden vahvistumista käännskielessä. Hypoteesin mukaan käännösten kielessä olisi siis piirteitä, jotka ovat tunnusmerkittäviä ja tavallisia kohdekielessä: käännökset suosisivat tyypillisiä syntaktisia rakenteita, välimerkitystä, kollokaatioita ja kliseitä. Lisäksi konventionaalisuuteen on liitetty myös oletus käännösten kieliopillisuudesta. (Baker 1996: 183–184.) Maurasen havainnot ovat ristiriidassa nimenomaan tyypillisiä kollokaatioita koskevan oletuksen kanssa. Leksikon konventionaalista kollokoitua vastaan puhuu osittain myös Kennyn (2001) tekemä tutkimus, jossa hän on tarkastellut hapax legomenoja eli vain kerran aineistossa esiintyviä sanoja ja kirjoittajaspesifisiä lekseemejä sekä näiden kollokoitua: tulokset osoittavat, että vaikka aineistossa onkin viitteitä ilmausten normalisoimisesta, useimmissa tapauksissa luovat ratkaisut (sanavalinnat ja kollokaatit) on jätetty normalisoimatta. Toisin kuin Maurasen tutkimus, Kennyn analyysi perustuu kuitenkin lähdetekstien ja käännösten vertailuun, joten tulokset eivät ole verrannollisia.

Lähdekielen vaikutus. Interferenssi lienee yksi useimmin esiin tulevista käsitteistä käännskisiin kohdistuvassa tutkimuksessa. Sillä tarkoitetaan kielessä ilmenevää häiriötilaa, joka on jonkin toisen kielen aiheuttamaa:

interference

A translation *error* that results from ignorance or a methodological *error* and that *introduces a characteristic peculiar to the source language* into the target text. (Humbley ym. 1999: 148; korostus JJ.)

interference

A term used in sociolinguistics and foreign-language learning to refer to the *errors* a speaker introduces into one language *as a result of his contact with another language*. (Crystal 1985: 160; korostus JJ.)

Määrityksistä on selvästi havaittavissa, että interferenssi ilmenee virheinä. Täten se on piirre, jota on kommunikaatiotilanteissa pidetty vältettävänä. Universaalitutkimuksen myötä interferenssiin on myös alettu suhtautua toisin: se on kääntämisessä ilmenevä lainalaisuus, jota ei voi eikä tarvitsekaan aina välttää. Interferenssi ei ole Bakerin alun pitäen luettelemien universaalihypoteesien joukossa, vaan yleisenä lainalaisuutena sitä on käsitellyt Toury (1995), jolta vuorostaan Laviosa (Laviosa-Braithwaite 1998) on poiminut sen universaalien joukkoon. Toury perustaa oletuksensa siihen, että kääntämisessä lähtötekstin piirteitä siirretään aina jossain määrin kohdetekstiin, ja mitä enemmän kääntäjä noudattaa lähdetekstin ilmaisua, sitä enemmän kohdetekstistä on löydettävissä merkkejä interferenssistä. Hänen mukaansa interferenssi ei ole kuitenkaan aina negatiivista (eli kohdesysteemin kodifioitujen käytänteiden rikkomista) vaan se voidaan myös nähdä positiivisena piirteenä, jolloin kyse ei ole säännöstöön kohdistuvista poikkeamista vaan distributiosta poikkeamisesta. Se, miten interferenssiä keitetään kohdesysteemissä,

vaihtelee ja riippuu lähdekulttuurista: Touryn mielestä prestiisin omaavista kulttuureista ja kielistä käännettäessä interferenssiä siedettäisiin etenkin silloin, kun kohdekulttuuri tai -kieli olisi prestiisiltään heikompaa. (Toury 1995: 275–278.) Mauranen (2000b: 136) puolestaan korostaa, että interferenssi olisi nähtävä laajempänä ilmiönä, jolloin sitä ei tarkasteltaisi lähde- ja kohdesysteemien välisenä ilmiönä, vaan yleisenä kääntämistä koskevana lainalaisuutena.

Lähdekielen vaikutusta on tarkasteltu universaalihypoteesina Eskolan (2002) tutkimuksessa, jossa sitä käsiteltiin neutraalina ja abstraktina käännöskielen piirteinä. Eskola huomauttaa, että interferenssin tutkimisessa on alettu yhä enemmän ottaa tutkimuskohteeksi lähde*kieli*, ei pelkästään siis lähde*teksti*, ja että tutkimus on vapautumassa preskriptiivisestä painolastista (mts. 264). Eskola saa tutkimukseensa tukea lähdekielen vaikutuksesta esitetyille universaalihypoteesille: tutkittuaan käännösten lähdekielten (venäjän ja englannin) systeemisiä mahdollisuuksia ja vastaavasti kohdekielen (suomen) systeemisiä piirteitä Eskola havaitsi, että lähdekielten systeemiset piirteet vaikuttavat tutkittujen syntaktisten piirteiden distribuutioon suomennoksissa (mts. 263–264). Myös aivan viimehetkinen tutkimus (Mauranen 2004a) on tuonut tukea Touryn esitykselle interferenssistä yleisenä lainalaisuutena.

3.3.2. Universaalitutkimuksen käsitteiden ja tutkimusmetodologian problematiikkaa

Edellä kuvatuilla universaalihypoteeseilla on siis tarkoitettu kielellisiä piirteitä, joita on käänöksissä riippumatta kielistä, joiden välillä kääntäminen on tapahtunut. Bakerin tunnetuksi tekemä universaali-käsite on kuitenkin monin tavoin ongelmallinen. Ensinnäkin hypoteesien tulkitseminen yleismaailmallisiksi kaikkea kääntämistä koskeviksi on tällä hetkellä mahdotonta, koska sekä hypoteesien perustana olevat havainnot samoin kuin tähän mennessä tehty tutkimus ovat keskittyneet pelkästään nykykieleen ja hyvin suppeaan kielivalikoimaan. Paloposki (2002) huomauttaakin, että universaali-käsite ei ole ottanut huomioon historiallista ulottuvuutta, joten oletuksia ei voi käsittää universaaleina piirteinä vaan tiettyä aikaa koskevinä säännönmukaisuuksina. Tymoczko (1998: 654) puolestaan huomauttaa, että pelkästään prototyyppisten käännösten³³ sisällyttäminen korpuksiin on ongelmallista universaalitutkimuksen kannalta. Tällöin jäävät ulkopuolelle sellaiset tekstit, jotka eivät ole prototyyppisiä (prototyyppisyys on tutkijan määrittelemää), jolloin hypoteesien testaaminen universaalina ilmiönä on mahdotonta. Chesterman (2004: 42) laajentaa tekstien valinnan koskemaan edustavuuden ongelmaa. Hän kysyykin, että jos tarkoituksena on tarkastella käännösten universaaleja ominaisuuksia, eikö tutkimuksen pitäisi silloin kattaa käännökset monipuolisesti – yhtä hyvin ammatti-

³³ Toury (2004: 20) huomauttaa, että prototyyppisen käännöksen määrittelemineen on sillään jo hyvin ongelmallista (ks. myös Halverson 1998).

laisten kuin ei-ammattilaistenkin käännökset ja myös huonoiksi käännöksiksi katsotut tekstit³⁴.

Kuten jo aiemmin tuli ilmi luvussa 3.2.2, nykyisten verrannollisten korpusten rakenne on universaalitutkimuksen näkökulmasta osittain puutteellinen sekä kieli- valikoimaltaan että genrevalikoimaltaan, joten tässä suhteessa yksikään korpus ei pysty vastaamaan universaalihypoteeseiksi esitettyihin väitteisiin. Varsin luultavaa on, ettei tällaiseen pysty yksikään tulevakaan korpus, sillä mikään tekstikokoelma ei pystyne toimimaan otoksena kaikesta tuotetusta käännösmateriaalista, saati sisältämään kaikkea olemassa olevaa tekstimateriaalia. Universaalihypoteesien testaus rajoittuukin nykykorpuksilla aina kerrallaan ajallisesti ja tekstivalikoimaltaan rajallisen materiaalin testaamiseen. Tähänastiselle tutkimukselle ja tulosten tulkinnalle ovat asettaneet rajansa myös aineiston kieli- valikoima, tekstien edustama kulttuuri, tarkasteltavan kääntäjän rooli ja esimerkiksi kustannustoiminnan merkitys (tarkemmin ks. Chesterman 2004: 41). Analyysistä saatuja tuloksia kokoamalla voidaan kuitenkin ryhtyä hakemaan laajempaa evidenssiä oletuksille ja toisaalta tarvittaessa muokkaamaan hypoteeseja. Tulosten vertailu voi tosin osoittautua vaikeaksi, mikäli korpusten koostamisperiaatteet ovat olleet hyvin erilaisia. Halverson (1998: 495) peräänkuuluttaakin korpusten koostamisperiaatteiden ja rakenteen sekä tutkimuskysymysten eksplisiittistä kuvausta eri tutkimuksista saatujen tulosten vertailun mahdollistamiseksi.

Jo tähän mennessä harjoitetun universaalihypoteesien testauksen aikana vaati- musta hypoteesien yleismaailmallisuudesta on lievennetty. Bernardinin ja Zanettin (2004: 52) mukaan universaalisuus on alettu käsittää yleiseksi tendenssioletukseksi sen sijaan, että se käsitettäisiin absoluuttiseksi totuudeksi. Tämä näkyy uutena käsitteenmäärittelynä esimerkiksi Eskolan (2002) tutkimuksessa: sen sijaan että hän testaisi universaaleja piirteitä, hän tarkastelee *lokaaleja lainalaisuuksia* ja peilaa omia tuloksiaan suhteessa muihin universaalihypoteeseja koskeviin tutkimuksiin. Näin ollen Eskola ei pyrikään testaamaan Laviosan (Laviosa-Braithwaite 1996) tavoin laajasti universaalihypoteeseja, vaan suhteuttamaan omat tuloksensa jo aiemmin saatuihin tuloksiin, mikä osoittaa selvää näkökulman muutosta. Tämänhetkiselällä korpuspohjaisella käännöstutkimuksella pystytäänkin tavoittamaan kerrallaan kapea osa universaali- tai lainalaisuushypoteesien kentästä.

Toinen seikka, jonka haluan nostaa esille, on universaalihypoteesien luomiseen liittyvä ongelma, johon olen viitannut jo aiemmin yksinkertaistumishypoteesia testatessani (Jantunen 2000, 2001a). Bakerin esittämien universaalioluetusten taustalla on runsas joukko erilaisia havaintoja, jotka perustuvat a) manuaaliseen vertailuun, b) pieneen aineistoon, jossa tarkastellaan muutamia tekstejä ja c) lähdeteksti–käännös-vertailuun. Näistä kaksi ensin mainittua eivät ole ongelmallisia siirryttäessä tarkastelemaan hypoteeseja, jotka on luotu alkuperäiskielen ja käännöskielen vertailuun korpusaineistolla. Manuaalisiin pieniin aineistoihin perustuvat hypoteesit

³⁴ Tirkkonen-Condit (henkilökohtaisesti 2003) on huomauttanut, että tutkimuksen tavoitteet ratkaisevat, millaisia tekstejä aineistoon on järkevää sisällyttää. Mikäli tavoitteena on tarkastella käännösprosessin aikaansaamia muutoksia kielessä, on kannattavaa tutkia mahdollisimman prototyyppisiä käännöksiä, jolloin mahdollisten häiriötekijöiden, kuten opiskelijakäännösten puutteellisesta ammattitaidosta johtuvat virheet, eivät pääse vaikuttamaan tuloksiin.

käyvät hyvin testattaviksi isoilla sähköisessä muodossa olevilla aineistoilla. Ongelmallista on sen sijaan hypoteesien nouseminen lähdeteksti–käännös-vertailusta, johon perustuvat havainnot ovat esimerkiksi Laviosan tutkimusten taustalla. Mielestäni ei voida aivan yksiselitteisesti olettaa, että se muutos tai vaihtelu, mikä tapahtuu lähtötekstin (tai -kielen) ja käännöksen (tai käännöskielen) välillä, olisi nähtävissä myös alkuperäiskielen ja käännöskielen välillä. Periaatteellisena ongelmana oletuksessa on se, että hypoteesien taustalla olevat havainnot perustuvat *kahta eri kieltä* edustavien tekstien väliseen vertailuun, kun taas hypoteesien lainalaisuuksien oletetaan koskevan *yhden kielen kahta eri muotoa* (ks. Jantunen 2001a: 99). Ei voitane olettaa suoraviivaisesti, että kahden käännösprosessissa mukana olevan kielisysteemin välillä tavattavat erot toteutuisivat samalla tavalla myös yhden kielen varianttien eli tässä tapauksessa käännöskielen ja alkuperäiskielen välillä. Andrew Chesterman (2001, 2004) onkin ehdottanut hypoteeseja jaettaviksi niin sanottuihin *S-* ja *T-universaaleihin*,³⁵ joista ensimmäiset koskevat lähde- ja kohdekieliä ja jälkimmäiset käännöksiä ja niille rinnasteisia alkuperäiskielisiä tekstejä. Chestermanin jaottelu selkiinntää käännöskielen tutkimuksen hypoteesien luomista, sillä sitä soveltamalla voidaan tehdä selvemmäksi se, mitä tutkitaan ja millaisilla aineistolla tutkimusta tehdään ja mitkä ovat oletusten taustalla olevat peruslähtökohdat.

Hypoteesien luomisen ongelmasta otan esimerkiksi hypoteesin kielen yksinkertaistumisesta. Hypoteesin taustalla olevat havainnot on tehty muun muassa hepreasta (Blum-Kulka ja Levenston 1983) ja unkarista (Klaudy 1996) englantiin tehdyistä käännöksistä, joissa oli käytetty yläkäsitteitä ja tuttuja synonyymeja, koska kohdekielessä ei ole alkuperäiskielen ilmauksille selkeää vastinetta. Kyse on siis Chestermanin jaottelussa *S-universaaleihin* liittyvästä havainnosta. Tällaista tilannetta ei luonnollisestikaan ole yhden kielen eri tavalla syntyneiden kielimuotojen välillä (*T-universaalit*): sekä käännös- että alkuperäiskielen ilmaisukeinoja eivätkä rajoita kielisysteemin eroista johtuvat seikat. Yksinkertaistumishypoteesin ja Laviosan (Laviosa-Braithwaite 1996) havaintojen käsittelyn yhteydessä olenkin huomauttanut (Jantunen 2001a: 99–100), että kahden eri kielisysteemin välisistä eroista (*S-universaalit*) johtuvaa ja niiden välillä havaittavaa leksikaalista yksinkertaistumista ei ole *välttämättä* havaittavissa yhden kielen eri muotojen välillä (*T-universaalit*). Yksi mahdollinen selitys tähän voi olla sanaston kontekstuaalinen merkitys, joka ainakin osittain tekee mahdottomaksi sen, että sanastosta valikoituisi vain osa käännöksiin (ks. myös mts. 101–103).

Kun Bakerin postuloimat hypoteesit perustuvat lähdetekstien ja käännösten vertailuun, myöhemmin esitetyt epätyypillisiä leksikaalisia kombinaatioita (Mauranen 2000b) ja uniikkeja aineksia (Tirkkonen-Conditi 2002, 2004) koskevat hypoteesit on luotu alkuperäis- ja käännöskielen vertailun pohjalta ja sopivat siten lähtökohdiltaan paremmin laajempaankin käännöskielen ja alkuperäiskielen vertailuun. Uusimmat hypoteesit ovat myös aiempia tarkkarajaisempia, mikä parantaa eri kielistä saatujen tutkimustulosten vertailtavuutta. Mauranen ja Tirkkonen-Conditin hypoteesit luokituvat Chestermanin *T-universaaleihin*. (Käännösuniversaalien tutkimusta esittelee laaja-alaisesti Mauranen – Kujamäki 2004.)

³⁵ S = source text, T = target text.

4. SYNONYMIA: KONTEKSTUAALISUUS JA KÄÄNNÖSKIELI

Synonymia kiinnostaa monia. Paitsi semantikot, siitä ovat kiinnostuneet niin kielen oppimisen kuin kääntämisenkin tutkijat, ja tavallinenkin kielenkäyttäjä innostuu varsin helposti pohtimaan jonkin ilmiön useanlaisia ilmiösujuja. Synonymian konseptuaalisen määrittämisen rinnalle on noussut erityisesti korpuslingvistiikan kehittymisen myötä myös kontekstuaalinen näkökulma, jossa samaa tarkoittavia ilmauksia tarkastellaan kontekstilähtöisesti. Seuraavassa tarkastelen aluksi synonymian erilaisia luonnehdintoja, minkä jälkeen keskityn synonymian kontekstuaaliseen luonteeseen ja erityisesti niihin kontekstuaalisiin piirteisiin, joiden on aiemmissa tutkimuksissa katsottu synonyymeja erottavan; tältä osin tarkastellaan jo joitakin edellä (lähinnä luvussa 2) käsiteltyjä esimerkkejä nimenomaan synonymien kontekstuaalisuuden kannalta. Viimeiseksi luon katsauksen käännettävyyden tutkimuksen piirissä käytyyn keskusteluun synonyymiasta.

4.1. Tietynasteista samuutta

Synonymia on yksi paradigmaattisen merkityssuhteen muodoista muun muassa hyponymian ja antonymian rinnalla. Yleensä sillä tarkoitetaan kahden tai useamman muodoltaan erilaisen lekseemin samanmerkityksisyyttä, mutta synonyymiksi voidaan käsittää myös lekseemiä laajemmat yksiköt kuten lausekkeet ja lauseet. Niinpä lausekkeet

- a) *kaunis nainen*
- b) *ulkoisilta ominaisuuksiltaan viehättävä täysikasvuinen naissukupuolta oleva ihminen*

voidaan tulkita toistensa synonyymeiksi, samoin kuin lauseet

- a) *Onpas täällä kylmä.*
- b) *Voitko sulkea ikkunan?*

jos näiden tilanteinen pragmaattinen merkityssisältö on sama eli niiden lausuja haluaa ikkunan suljettavaksi.

Samamerkityksisyydellä viitataan usein siihen, että ilmauksilla on identtinen tai riittävän samanlainen denotaatio eli ne rajaavat jostakin joukosta saman tarkoitteen. Siten esimerkiksi *ruoka* ja *sapuska* ovat toistensa synonyymeja, jos ne poimivat referenttikandidaattien joukosta saman tarkoitteen. Synonymia on siis kognitiivisen merkityksen samuutta: kielenkäyttäjät ymmärtävät ilmausten viittaavan samaan asiaan. (Kangasniemi 1997: 42; Karlsson 1999: 219; Kuiri 2000: 21–22.) On kuitenkin huomattava, että synonymia ei täyty tilanteisesta samaviitteisyydestä. Esi-

merkiksi *lahja* ja *kirja* voivat jossakin tilanteessa viitata samaan referenttiin eli lahjaksi saatuun tai annettavaan kirjaan, mutta samanlaisesta viittaussuhteesta huolimatta *lahjaa* ja *kirjaa* ei pidetä synonyymeina. (Kuiri 2000: 22.)

Cruse tähdentää, että synonyymien merkitykset ovat limittäisiä eli niillä on keskenään yhdistäviä jos myös erottaviakin piirteitä. Merkityksen samuus toteutuu synonyymeilla siinä, että ne ovat identtisiä *keskeisimmiltä* semanttisilta piirteiltään mutta voivat erota *perifeerisiltä* piirteiltään. (Cruse 1986: 266–267.) Jos keskeisenä yhdistävänä piirteenä pidetään viittaussuhteen samanlaisuutta, perifeerisiä erottavia ominaisuuksia ovat muun muassa konnotatiiviset merkitykset, joita synonyymisilla ilmauksilla on. Kuiri (2000: 22) toteaaakin, että juuri konnotatiiviset sivumerkitykset erottavat esimerkiksi sellaiset sanat kuin *akka* ja *nainen* toisistaan: synonyymin valinnalla heijastamme siis asenteitamme entiteettiä kohtaan. Lyons (1968: 427) kuitenkin huomauttaa, että emotionaaliset merkitykset voivat vaihdella eri puhujien välillä. Siten *akka* voi saada kahdella eri kielenpuhujalla erilaisia konnotatiivisia merkityksiä: siinä missä *akka* on jollekin ’naapurin pahansisuihin emäntä’, voi se jollekin olla ’oma rakas vaimo’.³⁶

Viittaussuhteeseen perustuva määrittely ei suinkaan ole ainoa keino määrittellä synonymiaa. Yhdeksi synonymian lajiksi on ehdotettu propositionaalista synonymiaa, joissa keskeisiä käsitteitä ovat totuusehto ja vaihdettavuus. Cruse (1986: 88, 2000: 158) määrittelee propositionaaliset synonyymit seuraavasti: x on y :n propositionaalinen synonyymi, jos (1) x ja y ovat syntaktisesti identtisiä ja (2) jokaisella kieliopillisella lauseella S , jossa on x , on samanlaiset totuusehdot kuin lauseella S_1 , jossa x on korvattu y :llä. Kyseessä on siis kahden ilmauksen samanlaisen totuusehdon täytyminen silloin, kun synonyymisia sanoja vaihdetaan ilmausten kesken. Niinpä *ruoka* ja *sapuska* ovat toistensa propositionaalisia synonyymeja ilmauksissa *Ruoka on valmista* ja *Sapuska on valmista*, koska molemmissa ilmauksissa sama totuusehto täytyy: syötäväksi tarkoitettu ravinto on valmista.

Propositionaalisetkaan synonyymit eivät ole kuitenkaan täysin identtisiä, kuten on havaittavissa *ruoka*- ja *sapuska*-sanoista. Totuusehdon täyttymisestä huolimatta niiden käyttöön liittyy tyyli- ja rekisterivalintoja sekä eroja ekspressiivisessä eli kielenkäyttäjän tunteita ilmaisevissa merkityksissä (Cruse 2000: 158). Sandersiin (1977) nojautuen Saukkonen (1984: 112) esittää synonyymien tyylillistä valintaa ohjaaviksi sekä kielenkäyttäjän että ympäristön tyylikonventioihin liittyvät tekijät (yksilöllis-sosiaalinen tyyli) ja tilanteiset, funktionaaliset ja tekstilajikohtaiset tekijät (kollektiivis-sosiaalinen tyyli). Esimerkiksi edellä mainittujen *ruoka*- ja *sapuska*-sanojen tyylilliseen valintaan vaikuttavat yhtäältä kielenkäyttäjän ja häntä ympäröivän sosiokulttuurisen ympäristön konventiot, ja toisaalta valintaa ohjaavat rajoitukset ja preferenssit, jotka ovat tyypillisiä tietyille tekstilajeille ja kielenkäyttötilanteille. Lisäksi niiden ekspressiivinen (konnotatiivinen) merkitys voi olla erilainen: *sapuska* ei ole välttämättä yhtä arvostettua kuin *ruoka*, ja lisäksi *sapuskalla* voidaan ilmaista erilaisia tunnetiloihin liittyviä ja tilanteisia merkityksiä.

Lähes aina synonymiaa pohdiskelevissa katsauksissa nousee esiin absoluuttisen synonymian ongelma. Tutkijoiden kesken vallitsee yksimielisyys siitä, että absoluuttista synonymiaa ei ole olemassa (mm. Lyons 1968: 447; Cruse 1986: 268, 270;

³⁶ Konnotatiiviset merkitykset ovat peräisin havaitsemistani todellista kielenkäyttötilanteista.

Kuiri 2000: 21), ja tätä on pidetty jo lähes itsestään selvyyttenä (jo Ullmann vuonna 1962, ks. Lyons 1968). Absoluuttisen synonymian mahdottomuutta on perusteltu kahdella seikalla: Ensinnäkin absoluuttinen synonymia olettaisi täyttää vastaavuutta denotatiivisessa ja propositionaalisessa merkityksessä sekä sivumerkityksissä että esiintymäympäristöissä, mikä on käytännössä mahdottomuus kahdelle ilmaukselle. Vaikka denotatiivinen merkitys olisikin kahdella tai useammalla ilmaisulla sama, niiden sivumerkityksissä ja esiintymäympäristöissä on yleensä eroja. Samoin on laita propositionaalisten synonyymien kanssa: vaikka totuusehdot olisivatkin samanlaisia, sivumerkitykset ja esiintymäympäristöjen erot erottavat nämäkin toisistaan.

Cruse vuorostaan korostaa ilmausten olevan absoluuttisia synonyymeja vain, jos niiden kontekstuaaliset suhteet ovat samanlaisia. Samanlaisuus koskee tässä tapauksessa sitä, että absoluuttisen synonyymiparin jäsenien on voitava esiintyä samoissa, kieliopillisesti mahdollisissa normaaleissa konteksteissa. Samoin esiintymisen outouden tai mahdottomuuden on koskettava molempien ehdokkaiden jäseniä. (Cruse 1986: 268, 2000: 157.) Käytännössä tämän toteaminen osoittautuu mahdottomaksi, koska kaikkia esiintymäympäristöjä on mahdotonta tavoittaa ja yksikin esiintymäympäristö, joka erottaa synonyymit toisistaan, riittää rikkomaan synonymian absoluuttisuuden. Toinen huomioon otettava seikka on kommunikation taloudellisuus: on nimittäin epätodennäköistä, että kielessä olisi kaksi aivan samaa merkitsevää ja funktioiltaan samanlaista ilmausta. Mikäli tällaisia syntyisi kieleen, synonyymisista ilmauksista toinen kuolisi pois motivaation puutteesta tai sen merkitys muuttuisi toisenlaiseksi. (Cruse 1986: 270; Partington 1998: 31.) Absoluuttinen synonymia -käsitteen arvo on lähinnä siinä, että se on eräänlainen oletus synonymian äärirajasta (Cruse 2000: 157). Jos toisena äärirajana on (mahdoton) absoluuttinen synonymia, on toisena kahden ilmauksen synonyymittömyys. Näiden ääripäiden välillä on jatkumo, jossa ilmausten synonyymisyys vaihtelee (Lyons 1968:447; Cruse 1986: 268).

Synonyymisten ilmausten täydellisen merkitys- ja funktiovastaavuuden mahdottomuuden vuoksi synonyymisina pidetään sellaisia ilmauksia, joilla on samanlainen konseptuaalinen tai propositionaalinen merkitys mutta joiden suhde ei muodosta täydellistä keskinäistä vaihdettavuutta kaikissa konteksteissa. Näin synonymian määrittelee esimerkiksi Carter (1987: 19). Tähän päättyy periaatteessa useampikin määritelmä, koska täydellinen vastaavuus on jouduttu hylkäämään. Absoluuttisen synonymian puuttumista on erikseen korostettu käsitteillä lähisyronyymit (Cruse 2000: 159–160) ja kvasisynonyymit (Leech 1981: 17). Ne ovat tosin redundantteja käsitteitä, koska, kuten todettu, absoluuttisen vastaavuuden puute on synonyymien inherentti ominaisuus; kaikki synonyymit ovat väistämättä lähisyronyymeja eivätkä täydellisiä. Eronteko on tässä mielessä turha, joskin lienee tarpeen silloin, kun vastaavuuden ongelmaa halutaan erityisesti korostaa.

Synonymiaa pidetään toisinaan vain itsenäisten sisältösanojen ominaisuutena. Tällaiseksi asiantilan tulkitsee Kuiri (2000: 22), joka rajaa kieliopilliset sanat synonymian ulkopuolelle: tarkastelleessaan *vuoksi-* ja *takia-*sanojen samankaltaisuutta hän päättyy siihen, että ”nämä partikkelit ovat siinä määrin kieliopillisiä elementtejä, ettei niitä voi pitää synonyymeina, koska synonymia on itsenäisten sanojen välinen suhde”. Rajanteko johtunee Kuirin synonymiamäärittelystä, joka nojaa ilmauksia

yhdistävään samanlaiseen denotaatioon – sellaistaahan ei kieliopillisilla sanoilla ole, koska ne eivät viittaa kielenulkoiseen maailmaan kuten sisältösanat. Mutta millainen *vuoksi*- ja *takia*-sanojen suhde sitten on, ellei synonyyminen, vaikkapa alla olevissa lauseissa?

- a) *Turha sen vuoksi on yöunia menettää.*
- b) *Turha sen takia on yöunia menettää.*

Eittämättä suhde on synonyyminen, vaikka denotaatio puuttuukin. Synonymia onkin tässä todettavissa propositionaaliseksi, sillä lauseilla on samanlainen totuusehto: jokin asiantila on niin turha, että yöunia ei kannata menettää. Vastaavasti Kangasniemi (1997: 42) pitää postpositioita *asti* ja *saakka* synonyymeina. Lyons (1968: 427) esittääkin, että ilmaukset voivat olla synonyymisia, vaikka niillä ei olisikaan viittaussuhdetta kielen ulkopuolisiin entiteetteihin. Huomattava on myös se, että edellisillä kahdella lauseella ei tunnu intuitiivisesti olevan emotionaalisia, tyyllisiä, sen enempää kuin rekisterikohtaisiakaan eroja.

Ylipäättään jakoa kieliopillisiin ja sisältösanoihin on kritisoinut mm. Leino (1993: 172–173; ks. myös 1990: 21), jonka mukaan jaottelu on perusteeton, koska kieliopilliset ainekset eivät ole pelkkiä merkitsimiä vaan niillä on myös oma itsenäinen merkityksensä. Tämän perusteella niitä ei voi rajata erilleen sisältösanoista. Jos myös kieliopillisilla sanoilla katsotaan olevan oma merkityksensä, eivätkö grammaattiset sanat, joilla on samanlainen merkitys – joskin ilman selvää denotaatiota – ole myös synonyymisessa suhteessa? Tästä käyvät edellisten lisäksi vielä esimerkiksi Onikki-Rantajääskön (2001: 227–234) käsittelemät paikallisjaiset asennonilmaukset *jaloillaan*, *seisaallaan*, *tolpillaan*, *pystyssä* ja *ylhäällä*, jotka voivat kaikki saada yhdistäväksi piirteeksi ihmisen vertikaalisen asennon kuvauksen ja jotka adverbeina kuuluisivat kieliopillisten sanojen ryhmään. Onikki-Rantajääskö (mts. 228) huomauttaa kuitenkin, että lähimerkityksisiä ilmauksia ei voisi pitää synonyymeina, koska niiden mielikuvaisuudessa on eroja; esimerkiksi *jaloillaan* ilmaus on tyyppillisin neutraaleissa konteksteissa, joissa korostuu jalkojen varassa oleminen ja mahdollisesti liike, *tolpillaan* taas on värittyneempi ja jopa humoristinen vertikaalisen asennon kuvaus, jossa ei tyyppillisesti ole liikkeen mahdollisuutta. Onikki-Rantajääskön mainitsemista piirteistä osa (esim. konnotaatiot) on kuitenkin juuri sellaisia kontekstuaalisista piirteistä, jotka tyyppillisesti erottelevat synonyymeja. Seuraava esimerkki paljastaakin, että sanoja voidaan käyttää myös synonyymisesti:

Hän pysyi jaloillaan ~ seisaallaan ~ tolpillaan ~ pystyssä ~ ?ylhäällä humalasta huolimatta.

Ainakin joidenkin perinteisesti kieliopillisten sanojen osalta (esim. adpositiot *takia* ja *vuoksi*, *asti* ja *saakka*) herääkin kysymys, voivatko ne olla jopa synonyymisempia kuin sisältösanat, joiden synonymiaan liittyy selkeitä tilanteisia ynnä muita eroja. Suhteita osoittavien kieliopillisten sanojen synonymia ja niiden synonyymisyyden suhde sisältösanojen suhteeseen vaatisi vielä selvitystä.

4.2. Kontekstuaalinen synonymia

Edellä synonymian määrittelyn yhteydessä tuli ilmi, että täydellisen synonymian vaatimuksena on synonyymikandidaattien keskinäinen vaihdettavuus identtisissä esiintymäympäristöissä. Todetuksi tuli myös se, että tällainen vaatimus on osoittautunut liian rajoittavaksi. Koska synonyymisten ilmausten sattumanvarainen vaihdettavuus ei ole mahdollista, on niiden kontekstuaalisissa valinnoissa oltava sellaisia piirteitä, jotka ohjaavat synonyymisten sanojen valikoitumista tiettyihin käyttöympäristöihin. Lyons (1968: 452) toteaaakin, että samoin kuin muiden ilmausten merkitykset, myös synonyymien merkitykset ovat kontekstisidonnaisia. Toisaalta synonyymiset – kuten muutkaan – ilmaukset eivät yleensä esiinny kahta kertaa normaaleissa kielenkäyttötilanteissa täysin identtisissä kontekstissa, joten niiden merkityskään ei toistune ilman pieniä muutoksia koskaan aivan samanlaisena. Tähän johtaa jo pelkästään se, että ilmauksen konseptuaalinen merkitys voi vaihdella jossain määrin puhujalähtöisesti, esimerkiksi *talo* voi merkitä yhdelle puhujalle toisenlaista rakennusta kuin toiselle. Toisaalta synonyymiset ilmaukset ovat keskenään erilaisia siinä, että niiden kontekstuaalinen käyttö on erilaista – joko kielenulkoisen kontekstin tai kontekstin tasolla.

Keskinäistä vaihdettavuutta tarkasteltaessa on puhuttu *normaalista vaihdettavuudesta*, jolla on tarkoitettu sitä, että ilmausten olisi oltava vaihdettavissa normaalissa kielenkäytössä (Lyons 1968: 451; Cruse 1986: 268–269). Kyse ei siis ole ensisijaisesti kieliopillisesta normaaliudesta (joskin se on tietysti edellytys ilmauksen mahdollisuudelle) vaan uusuksen normaaliudesta. Lyons (mp.) antaa esimerkeiksi synonyymeista ilmaukset

- a) *cow* ('lehmä')
- b) *mature female bovine animal* ('täysikasvuinen naaraspuolinen nautaeläin'),

joista jälkimmäinen voisi esiintyä sanakirjassa *cow*-sanana määrittelyinä. Ilmaukset ovat synonyymiastaan huolimatta kuitenkin hyvin erilaisia nimenomaan normaalin kielenkäytön näkökulmasta, sillä jälkimmäinen esiintyy pelkästään tietyissä funktioissa ja ensimmäinen on osa tavallista kielenkäyttöä. Myös Cruse (mp.) korostaa kontekstisidoksista merkitystä synonyymien erona: nimenomaan ilmauksen normaalisuus ja epänormaalius syntaktisesti kelvollisissa rakenteissa erottaa synonyymikandidaatit toisistaan. Vaihdettavuutta käsiteltäessä puhutaan siis ensisijaisesti uusuksesta ja kontekstuaalisesta merkityksestä. Keskinäistä vaihdettavuutta on testannut laajasti Olkinuora (1992) synonyymisten *aalto-* ja *laine-*lekseemien avulla (samoja lekseemejä on käsitellyt myös Kangasniemi 1997: 42). Crusea (1986) mukaillen hän jakaa synonyymiehdokkaat sisältävät lauseet ”normaalimpiin” ja ”vähemmän normaaleihin” oman intuiionsa perusteella. Esimerkiksi lauseista

- a) *Suomenlahden aallot pauhasivat.*
- b) *Suomenlahden laineet pauhasivat.*

ylempi olisi Olkinuoran mukaan normaalimpi. Menetelmä toimii erityisesti tutkittaessa synonyymisten sanojen valikoitumista idiomeihin ja metaforisiin ilmauksiin. Vaihdeavuustesti paljastaa herkästi, että vain jokin synonyymisista lekseemeistä on vakiintunut tiettyyn kuvaannolliseen ilmaukseen tai idiomiin. Esimerkkejä Olkinuoralta (mts. 117):

- a) *Pommikoneitten ensimmäinen **aalto** saapui Helsinkiin* (normaalimpi).
- b) *Pommikoneitten ensimmäinen **laine** saapui Helsinkiin* (vähemmän normaali).

- a) *kuin lastu **laineilla*** (normaalimpi)
- b) *kuin lastu **alloilla*** (vähemmän normaali).

Kotekstiltaan identtisillä esimerkkilauseilla pystytään intuitioon perustuen testaamaan etenkin metaforissa ja idiomeissa esiintyvien synonyymien vastaavuutta. Vaihdeavuustestauksessa käytetyt lauseet eivät kerro kuitenkaan laajemmin *uusuksen* tyypillisyydestä, koska ne ovat yksittäistapauksia ja valikoituja esimerkkejä halutusta ilmiöstä. Vaikka intuitio pitäisikin molempia synonyymeja vapaasti vaihdettavina (normaaleina) tietyissä lauseissa, niiden ko(n)tekstuaalinen uusutus saattaa olla hyvinkin erilainen. Tätä voidaan testata ainoastaan laajoilla todellista kielenkäyttöä kuvaavilla aineistoilla, mikä paljastaa ilmausten kontekstuaaliset säännömukaiset preferenssit ja yleisen tason vaihdettavuuden.

Ilmausten kontekstuaalinen sidoksisuus on juuri se piirre, johon tavallinen kielenkäyttäjä yleensä kiinnittää huomiota pohtiessaan samaa merkitsevien sanojen käyttöä: yhtäällä on *itikat* ja *vihdat*, toisaalla *lehmät* ja *vastat*. Synonyymisten ilmausten murrevaihtelun tarkastelu on ollut tavallista kielitieteessä ja erityisesti silloin, kun sanaston käytön eroja on popularisoitu (ks. esim. Länsimäki 1995; Puntila 1995). Sekä alueellinen että myös sosiaalinen variaatio kuuluvat ilmausten käyttösääntöjen uloimpaan kerrokseen, kontekstuaaliseen merkitykseen, yhdessä muun muassa tyylipiirteiden ja rekisterikohtaisessa käytössä ilmenevien rajoitusten ja preferenssien kera. Tällaiset vakiintuneet konventiot koskevat luonnollisesti myös synonyymeja ja toimivat nimenomaan samaa merkitseviä ilmauksia erottamassa (ks. mm. Cruse 1986: 283, Warren 1988: 79–80).

Synonyymisten ilmausten käyttöä koskevaa kaikkea kontekstuaalista tietoa ei ole mahdollista tavoittaa, koska ilmausten käyttöyhteyksiä on lukuisia. Siksi on tyydyttävä jossakin määrin likimääräistyksiin. Synonyymien *kotekstuaalisiin* suhteisiin keskittyvää tutkimusta on harjoitettu jossain määrin etenkin englannin kielestä. Tarkastelun kohteena on silloin ollut tyypillisesti synonyymisia ilmauksia ympäröivä sanasto eli synonyymien kollokointi. Tunnetuin lienee Hallidayn jo luvussa 2.3.1 mainittu havainto englannin STRONG- ja POWERFUL-adjektiivien ('voimakas') kollokoinnista. Siinä missä *strong tea* ja *a powerful car* ovat tyypillisiä ja hyväksyttäviä kombinaatioita, *powerful tea* ja *a strong car* eivät ole; synonyymit eivät ole siis keskenään vaihdettavissa tässä kontekstissa (Halliday 1966: 150). Myös luvussa 2.1 mainittu Leechin (1981: 17) englannin PRETTY- ja HANDSOME-adjektiiveja koskeva esimerkki osoittaa synonyymisille ilmauksille omanlaisensa kollokatiivisen valinnan.

Vastaavasti Persson (1989, 1990) on synonymiotutkimuksessaan tarkastellut englannin DEEP- ja PROFOUND-adjektiiveja (’syvä’), jotka eroavat paitsi kollokaateiltaan myös monilta muilta ominaisuuksiltaan:

Taulukko 4.1. DEEP- ja PROFOUND-adjektiiveja erottavia tekijöitä (Persson 1989, 1990).

	DEEP	PROFOUND
metaforinen status	positionaalinen (<i>He was deep in thought</i>) laadullinen (<i>deep learning</i>)	vain laadullinen (<i>profound learning</i>)
emotionaalisuus vs. intellektuaalisuus	emotionaalinen konteksti (<i>deep feeling, regret</i>)	intellektuaalinen konteksti (<i>profound capacity</i>)
intensiivisyys	heikko	voimakas
tyyli	epämuodollinen viestintä	muodollinen viestintä

Havaitut erottavat piirteet ovat osittain seurauksia toisistaan, sillä epämuodolliselle tyylille ovat tyypillisempiä emotionaalisuutta ilmaisevat kuin intellektuaalisuuteen viittaavat ilmaukset. Samoin on laita metaforisten ilmausten kanssa. Perssonin listaamat erottavat piirteet vaikuttavat myös adjektiivien kollokationaaliseen merkitykseen, sillä niitä ympäröivä leksikko jakautuu luonnollisesti samojen kontekstuaalisten periaatteiden mukaisesti. (Englannin kieltä koskevia havaintoja lisäksi mm. julkaisuissa Stubbs 1995a; Krishnamurthy 1996; Biber ym. 1998.)

Synonyymien kontekstuaalisuutta esittelevät havainnot eivät rajoitu kuitenkaan vain englannin kieleen, vaan ilmiötä on tarkasteltu suomen kielestäkin. Mainittakoon ensin Kangasniemen (1997: 42) esimerkki *iso-* ja *suuri-*adjektiivien käytöstä: vaikka ilmaukset *iso kivi* ja *suuri kivi* ovat samamerkityksisiä, *iso mies* ja *suuri mies* eivät välttämättä ole. Tämä on seurausta siitä, että *suuri* esiintyy tyypillisemmin abstrakteissa asiayhteyksissä ja *iso* vuorostaan konkreettisissa (vrt. *suurmies* vs. **isomies* vaikkakin *isokenkäinen*). Kangasniemen analyysi ei käsittäkseni perustu korpusanalyysiin, mutta sellaisiakin on tarjolla suomalaisista synonyymeista: olen aiemmin toisessa yhteydessä (Jantunen 2001b) tarkastellut kahden adjektiivin – TÄRKEÄN ja KESKEISEN – synonymiaa korpusmetodein. (Tätä esimerkkiä samoin kuin seuraavia Arppen tutkimuksia on käsitelty lyhyesti aiemmin luvussa 2 kontekstuaalisten piirteiden kuvauksen näkökulmasta). Tutkimuksessa synonymiaa erottaviksi piirteiksi erottui edellisten tutkimusten tapaan kollokointi mutta myös kolligointi, morfologinen preferenssi ja syntaktinen funktio. Kollokaatioanalyysissa tuli esiin synonyymien selkeä preferenssi tiettyihin kollokaatteihin sekä oikean- että vasemmanpuoleisissa kontekstissa. Näistä erottavista kollokaateista olen luetellut viisi merkitsevintä taulukoihin 4.2 ja 4.3.

Adjektiiveilla oli omanlaisensa kollokointi: erityisesti TÄRKEÄ-adjektiivin vasemmanpuoleisina kollokaateina olevat astemääritteet (*vähemmän*, *erityisen*, *yhtä*, *hyvin*) erottivat adjektiivit toisistaan, mutta myös muunlaista kollokationaalista

Taulukko 4.2. Synonyymeja TÄRKEÄ ja KESKEINEN erottavia oikeanpuoleisia vieruskollokaatteja (Jantunen 2001b).

TÄRKEÄ	KESKEINEN
SEIKKA	TEESI
YHTEISKUNNALLINEN <i>erityisesti</i>	ELEMENTTI
SUHDE	TEEMA
<i>kuin</i>	IDEA
	OSUUS

Taulukko 4.3. Synonyymeja TÄRKEÄ ja KESKEINEN erottavia vasemmanpuoleisia vieruskollokaatteja (Jantunen 2001b).

TÄRKEÄ	KESKEINEN
<i>vähemmän</i>	KAKSI
<i>erityisen</i>	POLITIIKKA
<i>yhtä</i>	TEORIA
<i>paljon</i>	TUTKIMUS
<i>hyvin</i>	TOINEN

variaatiota esiintyi. Adjektiivit erosivat myös sen mukaan, kummalle puolelle kollokointi oli vahvempaa: TÄRKEÄ kollokoi voimakkaammin vasemmanpuoleisen kontekstin kanssa, KESKEINEN taas oikeanpuoleisen.

Taulukossa 4.3 näkyvä TÄRKEÄ-adjektiivin kollokointi astemääritteiden kanssa heijastui voimakkaasti myös sanaluokkakolligaatiojakaumaan, sillä astemääritteiden vuoksi sen adverbikolligaatin osuus oli KESKEISTÄ huomattavasti suurempi vasemmanpuoleisten vieruskolligaatiensa joukossa (esim. *Ajallisen jatkuvuuden (samuuden) ohella myös erottautuminen on hyvin tärkeää*) ja osoittautui siten lähisynonyymeja erottavaksi tekijäksi. Oikeanpuoleisissa vieruskolligaatioissa oli myös eroja: esimerkiksi substantiivikolligaatin osuus oli KESKEINEN-adjektiivilla peräti noin 80 prosenttia kaikista kolligaaateista tässä asemassa, sen synonyymilla osuus oli huomattavasti pienempi, vain noin 47 prosenttia (esim. *Semioottisina oliona tekstejä yhdistää muuan tuiki tärkeä seikka* --). Synonyymien kolligaatioanalyysit ovat tietoni mukaan harvinaisia, mutta oman tutkimukseni lisäksi ainakin Biberin ym. (1998) aiempi korpusanalyysi englannin START- ja BEGIN-verbien kieliopillisista preferensseistä on luettavissa tällaiseksi. Tarkastellessaan verbien kieliopillisia yhteisesiintymiä he havaitsivat, että verbien objektikseen saamien *to*-lauseiden osuus vaihteli selvästi. Aineiston mukaan BEGIN saa *to*-lauseen objektikseen (eli kolligaa-tikseen; esim. *Then I began to laugh a bit*) lähisynonyymiaan huomattavasti useammin: sen osuus oli BEGIN-verbillä (sen mukaan, mikä tekstilaji on kyseessä) 36–60 %, START-verbillä vain 14–17 % (tarkemmin kyseisten synonyymien käytöstä mts. 95–100).

Olen edellä luvussa 2.3.5 postuloinut morfologisen preferenssin yhdeksi kontekstuaaliseen merkitykseen kuuluvaksi piirteeksi. Postulaation yhtenä perusteena oli nimenomaan synonyymeja TÄRKEÄ ja KESKEINEN analysoitaessa esiin tullut preferenssi tietynlaisiin muotoihin. Tarkasteltujen adjektiivien lekseeminsäisen morfologisen preferenssin ero tuli esille sekä tarkasteltaessa vertailuasteita että sijaituvutusta. Kontekstuaalisen morfologisen preferenssin ero näkyi taas V1-sub-

stantiivikollokaattien sijataivutuksessa genetiivien, partitiivien ja elatiivien määrän eroina. Tulokset on kuvattu tiivistäen taulukossa 4.4:

Taulukko 4.4. Synonyymien TÄRKEÄ ja KESKEINEN morfologinen preferenssi (Jantunen 2001b).

	TÄRKEÄ	KESKEINEN
<i>Paradigmaattinen:</i>		
nominatiivissa	33 %	43 %
partitiivissa	40 %	19 %
positiivimuotoinen	71 %	91 %
komparatiivimuotoinen	9 %	2 %
superlatiivimuotoinen	20 %	7 %
<i>Kontekstuaalinen:</i>		
V1: genetiivissä	46 %	70 %
V1: partitiivissa	14 %	2 %
V1: elatiivissa	14 %	5 %

Lekseeminsisäiselle preferenssille löytyi selitys tarkastelemalla adjektiivien syntaktista funktioita, jossa oli selkeitä eroja: KESKEINEN-adjektiivia oli käytetty ylivoimaisesti eniten adjektiiviattribuuttina nominatiivimuotoisen pääsanana kanssa, mutta sen synonyymia vuorostaan sekä attribuuttina että myös predikatiivina, mikä lisää partitiivin määrää.

Myös kahdessa muussa suomalaisessa tutkimuksessa on katsastettu morfologisia piirteitä synonymiaa erottamassa. Tutkiessaan verbejä PYSTYÄ ja VOIDA Arppe (2001) huomasi, että PYSTYÄ preferoi imperfekti- ja partisiippimuotoja ja VOIDA vuorostaan passiivia ja konditionaalia. Toisaalla (Arppe 2002) hän on havainnut myös variaatiota toisten synonyymien verbien, MIETTÄÄ ja POHTIA, välillä: kun MIETTÄÄ esiintyy tyypillisimmin yksikön 1. persoonassa, POHTIA preferoi sekä yksikön ja monikon 3. persoonaa vain muutamia havaintoja mainitakseni.

Synonyymien semanttista valintaa (semanttinen preferenssi ja prosodia) koskeva tutkimus on ollut olematonta. Satunnaisia havaintoja kontekstin semanttisista piirteistä tavataan kylläkin joissakin kollokaatioanalyseissa. Esimerkiksi Biber ym. (1998: 43–48) raportoivat englannin LARGE-adjektiivin suosivan pääsanoinaan sanoja, joiden merkitys on määrää ilmoittava, kuten NUMBER, AMOUNT, PROPORTION, SUM, mutta sen synonyymi BIG vuorostaan määrittää sellaisia sanoja, jotka ovat kooltaan fyysisesti mitattavissa, kuten MAN, HOUSE, TOE, BOY, ROOM. Kyseessä on kollokaatteja yhdistävien merkityspiirteiden analyysi eli semanttinen preferenssi, vaikkakaan tutkimus ei pyri tyhjentävään merkityspiirteiden analyysiin vaan tulokset ovat tutkimusta lukevan poimittavissa, mikäli tämä tietää mitä etsiä. Semanttista prosodiaa esitellessään Stubbs (1995a) on tarkastellut semanttisesti läheisiä lekseemejä, joista ainakin REASON, EFFECT ja CONSEQUENCE ovat jonkinasteisia synonyymeja. Näistä REASON on semanttiselta prosodialtaan neutraali, EFFECT taas tyypillisimmin ja CONSEQUENCE vahvasti negatiivinen (mts. 45–48). EFFECT-lekseemillä on myös muista poikkeava semanttinen preferenssi: sen kollokaateissa on usein lääketieteeseen liittyviä lekseemejä kuten DRUGS, PLACEBO ja PSYCHOLOGICAL. Myös Perssonin (taulukko 4.1 s. 59) tekemät havainnot liittyvät

osittain kotekstin semanttisiin piirteisiin, esimerkiksi DEEP-adjektiivin yhteisesiintyminen emotionaalisuutta ilmaisevien kollokaattien kanssa. Varsinaiset synonyymien semanttisten valintapiirteiden analyysiin keskittyvät tutkimukset ovat nähdäkseni kuitenkin vielä tekemättä; edellä mainitun kaltaiset tutkimukset kuitenkin antavat viitteitä siitä, että myös kotekstin semanttiset piirteet erottelevat merkitykseltään samankaltaisia ilmauksia toisistaan.

Aikaisemmissa tutkimuksissa on siis saatu osoituksia siitä, että synonyymiset lekseemit poikkeavat laajemman kontekstin lisäksi myös lähikotekstiltaan ja jopa taivutusmuodoiltaan. Tällaiset tulokset selittävät osaltaan absoluuttisen synonyymian mahdottomuutta. Jos merkitykseltään lähekkäiset sanat saattavat käyttäytyä kontekstuaalisesti ja taivutukseltaan eri tavalla, ei niiden keskinäinen vaihdettavuus ole silloin mahdollista. Tämä ei tietystikään sulje pois sitä mahdollisuutta, että jotkut synonyymit voivat olla kontekstuaalisesti lähempänä toisiaan kuin toiset. Kotekstuaalinen ja taivutusmuotoihin liittyvä preferenssi lienee myös jonkinlainen jatkumo, jonka toisessa päässä ovat suhteellisen vapaasti vaihdettavat synonyymit ja toisessa hyvinkin tarkasti tiettyjä rajoituksia noudattavat ilmaisut – kuten on laita kontekstuaalisessa valinnassa yleensäkin. Synonyymien käyttöä ohjaavat siis edellä kerrotun mukaisesti kontekstuaaliset valintaominaisuudet, jotka myös osoittautuvat synonyymeja erottaviksi tekijöiksi. Erikoistuminen tietynlaisiin kontekstuaalisiin käytötapoihin luo edellytykset synonyymien säilymiselle. Leksikossa ei siten liene joutomiehiä, jotka voisivat karsiutua pois suoranaisesti sen tähden, että niiden kontekstuaalinen valinta olisi identtinen. Muutoksen syyt lienevät muualla.

Käsillä olevassa työssä synonyymian toteutumana pidetään sellaisten ilmausten suhdetta, jossa ilmauksilla on samanlainen funktionaalinen (pragmaattinen) merkityssisältö ja keskenään samankaltainen denotaatio; jälkimmäinen vaatimus ei ole kuitenkaan ehdoton vaan synonyymisessa suhteessa oleviksi lasketaan myös kielipilliset sanat, joilla ei katsota olevan denotaatiota. Synonyymien katsotaan myös eroavan toisistaan cruselaisittain perifeerisiltä (kuten konnotaatioltaan) ja kontekstuaalisilta ominaisuuksiltaan.

4.3. Synonymia ja käännöskieli

Synonyymiset ilmaukset voivat olla ongelmallisia kääntämisessä niin kuin muissakin kielenkäyttötilanteissa, joissa on mukana kaksi kieltä tai kieli, joka ei ole yhtä vahva kuin oma äidinkieli. Jälkimmäinen on tuttua vieraan kielen oppimisesta. Käännettäessä synonymia voi koitua ongelmaksi ainakin kahdella tavalla: a) miten valita kohdekielen synonyymijärjestelmästä oikea vastine lähdekielen ilmaukselle ja b) mikä on synonyymipesyestä valitun lekseemin kontekstuaalinen valinta? Näistä ensimmäinen liittyy kahden eri kielen leksikon merkityspiirteiden tuntemukseen, jälkimmäinen taas kohdekielen järjestelmän tuttuuteen. Ongelmaksi kahden kielen välillä toimiessa voi muodostua esimerkiksi Bakerin (1992: 23–24) mainitsema ero sanojen ekspressiivisessä merkityksessä. Vaikka kahdella eri kielen lekseemillä olisikin samanlainen propositionaalinen merkitys, saattavat niiden muut merki-

tysominaisuudet, kuten juuri ekspressiivinen merkitys, poiketa kielten välillä. Tällaisia ongelmia voi Bakerin mukaan tulla muun muassa uskontoon, politiikkaan ja seksuaalisuuteen liittyvässä sanastossa, jotka ovat voimakkaasti kulttuurisidoksia.

Synonyymien kontekstuaalisidoksisuus voi vuorostaan muodostua ongelmaksi silloin kun kääntäjä tekee työtään vieraaseen kieleen päin. Samaan tapaan kuin kieltenoppijakin hän voi törmätä vaikeuksiin valita rekisteriin, tyyliin tai kontekstiin sopiva ilmaus erilaisista samaa merkitsevistä vaihtoehdoista. Vieraan kielen käyttäjä ei aina pysty natiivin kielenpuhujan tavoin valitsemaan kielitietonsa avulla oikeaa synonyymia, jolloin hän tekee uususta rikkovan virheen. Partingtonin mukaan tällaisia ongelmia voivat aiheuttaa erityisesti puolikieliopilliset sanat, kuten astemääritteet. Hän antaa esimerkin englannin astemääritteistä *rather* ja *somewhat*, joiden merkitys on sanakirjojen mukaan samanlainen ('jokseenkin') mutta joilla on kuitenkin eroja käyttötiheydessä ja kontekstuaalisissa valinnoissa (1998: 32, ks. myös mts. 53–). Tämänkaltainen tilanne tuli vastaan omassa suomi vieraana kielenä -opetuksessani: opiskelija valitsi sanakirjansa avulla vahventavan astemääritteen määrittämään kvanttoria lauseeseen *Minulla on x paljon rahaa*, jolloin hän päätyi ratkaisuun *Minulla on aivan paljon rahaa*. Tällaisissa tapauksissa korpuksen perustuvat konkordanssianalyysit voivat auttaa vieraan kielen oppijaa sekä myös kääntäjää tämän kääntäessä vieraaseen kieleen päin. (Ks. Pearson 1996 ja Lindquist 1999 korpuksista kääntäjän apuna sekä Malmkjær 1993 kääntämisen ja L2-kielen oppimisen yhteisistä ongelmista.)

Käännettyä kieltä ja synonymian käyttöä koskevia havaintojakin on olemassa jonkin verran. Käännöskielen universaalitutkimusta tarkasteltaessa (luku 3.3) tuli mainituksi jo Blum-Kulkan ja Levenstonin tutkimus (1983) ja oma tutkimukseni (Jantunen 2001a), joissa on käsitelty synonyymiaa käännettyssä kielessä. Näiden lisäksi ovat havaintoja tehneet muun muassa Klaudy (1996) ja Mauranen (2000b). Tutkimuksista kolme ensiksi mainittua liittyy oletukseen käännöskielen yksinkertaistumisesta. Blum-Kulka ja Levenston (1983: 130–131) esittävät aiempiin havaintoihin nojautuen, että käännöskielessä on tyypillistä käyttää synonyymipesyeen frekventeimpiä ja tutuimpia synonyymeja harvinaisten ilmausten sijaan. Laviosan (Laviosa-Braithwaite 1996: 2) mukaan heidän havaintonsa perustuvat kuitenkin sellaisiin raamatunkäännöksiin, joissa tekstiä on tietoisesti muiltakin osin yksinkertaistettu erityistä lukijakuntaa varten, joten havaintoa ei hänen mukaansa voi yksiselitteisesti pitää todisteena käännöskielen leksikaalisesta yksinkertaistumisesta. Klaudy (1996: 150–151) on puolestaan havainnut, että unkarissa aamupäivää tarkoittavien sanojen kirjo (*hajnal* 'aikainen aamu', *reggel* 'aamu' ja *délelött* 'tunnit ennen keskipäivää') on hävinnyt käännettäessä indoeurooppalaisiin kieliin, joissa kaikkien vastineeksi on tullut frekventti 'aamua' merkitsevä sana, esimerkiksi saksan *morgen* ja englannin *morning*, sen sijaan että olisi käytetty ilmauksia *dawn*, *Tagesanbruch* ja *forenoon*, *Vormittag*. Viimein Kohn (1996: 48) on ehdottanut, että oletettu käännöskielen leksikaalinen yksinkertaistuminen voisi johtua nimenomaan synonyymien käytön yksinkertaistumisesta.

Oma tutkimukseni (Jantunen 2000, 2001a), jossa tarkastelin käännöskielessä käytettyjen synonyymien yksinkertaistumista, osoittaa kuitenkin päinvastaista. Tutkimukseni kohdistui astemääritteiden, tarkemmin maksimoijien (kuten *aivan*, *ihan*) ja vahventajien (esim. *hyvin*, *kovin*) distribuutioon akateemisten tekstien ja

kaunokirjallisuuden korpuksissa, jotka olivat peräisin Käännösuomen korpuksesta. Näiden synonyymien käyttöä koskeva analyysi osoitti, että käännettyssä kaunokirjallisuudessa ja akateemisissa teksteissä käytettyjen astemääritteiden kirjo oli yleensä yhtä suuri tai jopa suurempi kuin verrannollisissa alkuperäissuomalaisissa teksteissä. Testaus ei siis leksikaalisen variaation osalta tuottanut tukea yksinkertaistumiselle. Tutkimuksessa tarkastelin lisäksi sitä, onko käänöksissä käytetty frekventtejä synonyymipesyeen jäseniä enemmän kuin vastaavissa alkuperäissuomalaisissa teksteissä. Tätäkään tukevia tuloksia ei tullut ilmi. Tämä tapaustutkimus, joka aiemmista havainnoista poiketen vertasi yhden kielen kahta eri muotoa, ei tuonut tukea sille, että synonyymisia ilmauksia käytettäisiin käänöksissä eri tavalla kuin alkuperäiskielessä, eikä myöskään oletukselle leksikaalisesta yksinkertaistumisesta. Vastaavanlaista tutkimusta kaivattaisiin kuitenkin muunkinlaisilla synonyymiryppäillä, jotta synonyymien käytön yksinkertaistumista voitaisiin arvioida laajemmin.

Vastaavantyyppisestä Maurasen (2000b) korpusanalyysistä, jossa keskitytään metatekstuaalisten elementtien kollokointiin käänös- ja alkuperäiskielessä, tulee esiin mielenkiintoisia huomioita myös synonyymien käytöstä. Mauranen (mts. 125) havaitsee akateemisista teksteistä, että synonyymiparista *ensiksi–ensinnäkin*³⁷ jälkimmäinen on tyypillisempi alkuperäissuomalaisissa teksteissä ja ensimmäinen hieman tavallisempi käänöksissä. Vastaavanlaista variaatiota on YRITTÄÄ–PYRKIÄ-parilla: ensimmäinen on tavallisempi käänöksissä, jälkimmäinen taas alkuperäissuomessa. Parit osoittavat siis kielimuoto kohtaista vaihtelua. Frekvenssierojen lisäksi tutkimuksesta käy ilmi, että synonyymisten ilmausten kollokatiiviset suhteet voivat olla erilaisia kielimuodoissa. Verbeistä *korostaa* ja *painottaa* edellinen kollokoi HALUTA-verbin kanssa molemmissa kielimuodoissa, joskin se oli selvästi yleisempi alkuperäissuomessa, mutta HALUTA <*painottaa*> osoittautui selvästi käänösuomen piirteeksi. Maurasen tutkimus tuo synonyymisten ilmausten käytöstä siis esille eroja, jotka osoittavat paitsi erilaista preferenssiä synonyymien valinnassa myös erilaista kontekstuaalista preferointia alkuperäiskielen ja käänöskielen välillä. Mauranen (2000a) esittääkin, että käänöksistä on löydettävissä sellaisia alkuperäissuomelle vieraita leksikaalisia myötäesiintymiä, jotka esittelevät uusia synonyymisten ilmausten käyttötapoja.

³⁷ Kuiri (2000: 22) ei pidä samamerkityksisten samakantaisten johdosten suhdetta synonyymisenä, koska niiden muoto ei ole riittävän erilainen.

5. TUTKIMUKSEN TOTEUTTAMINEN: HYPOTEESIT, AINEISTO JA METODIT

Tämä jakso esittelee lähemmin tutkimuksen hypoteesit, aineistona olevan korpuk- sen ja sen koostamisperiaatteet ja työssä käytetyt menetelmät. Myös hypoteesien testausta varten valitut havaintoyksiköt ja niiden valinnan perusteet, vertailu- ja tilastolliset menetelmät sekä analyysissä käytetyt termit ja tarkastelualueen määrittely kuvataan tässä luvussa.

5.1. Tutkimushypoteesit

Tutkimukseni perustavoitteina on selvittää, miten kontekstuaalinen valinta erottaa synonyymit toisistaan ja miten käännösuomen leksikon kontekstuaaliset profiilit eroavat alkuperäissuomen profiileista. Tutkimustani varten asetan seuraavat hypoteesit, jotka on muodostettu luvuissa 2, 3 ja 4 esitettyjen synonymiaa, ilmausten kontekstuaalisuutta ja käännöskieltä koskevien aiempien tutkimusten ja teoriataustan pohjalta:

1. Synonyymisiksi luettuja sanoja erottavat toisistaan kontekstista löydettävät valintapiirteet, eli niiden kontekstuaaliset profiilit ovat erilaisia. Erottavia piirteitä on löydettävissä kaikista kontekstuaalisista merkityspiirteistä (kollokatio- naalisesta ja kolligationaalisesta valinnasta, morfologisesta preferenssistä sekä semanttisesta preferenssistä ja prosodiasta).
2. Käännösuomen sanaston kontekstuaaliset profiilit poikkeavat alkuperäis- suomesta tavtavista profiileista, eli käännösuomessa on sellaisia kontekstu- aalisia valintoja, jotka ovat alkuperäissuomelle epätyypillisiä.

Käännösuomea koskevat lisäksi vielä seuraavat sivuhypoteesit. Suluissa on mainittu, mihin luvussa 3.3.1 käsiteltyyn universaalihypoteesiin oletus liittyy.

- 2.1. Käännösuomesta on löydettävissä alkuperäissuomelle epätyypillisiä leksikaalisia ja kontekstuaalisia frekvenssejä (hypoteesi epätyypillisistä frek- vensseistä).
- 2.2. Käännösuomen leksikon kontekstuaalinen valinta eivät ole yksinkertaistunutta (yksinkertaistumishypoteesi).
- 2.3. Synonyymien käyttö ei ole kaventunutta käännösuomessa (yksinkertaistumishypoteesi).
- 2.4. Käännösuomen leksikaalisiin profiileihin vaikuttaa käännösten lähdekieli eli yhdestä lähdekielestä käännetyn suomen kontekstuaaliset profiilit ovat erilaisia kuin käännösuomessa yleensä (hypoteesi lähdekielen vaikutuksesta).

5.2. Tutkimusaineistona Verrannollinen kaunokirjallisuuden korpus

Tutkimuksen aineisto on peräisin Joensuun yliopiston kansainvälisen viestinnän laitoksessa kootusta Käännössuomen korpuksesta (KSK; Mauranen 1998a, 2000b), joka esiteltiin aiemmin luvussa 3.2.1. Aineiston koko ja laatu ovat määräytyneet seuraavassa esitettävien koontiperiaatteiden ja tutkimusasettelun mukaan. Tutkimuksessa on pyritty ottamaan huomioon korpuksen kattavuutta koskevat yleiset koontiperiaatteet KSK:n asettamissa rajoissa. Aineisto on koottu tähän työhön KSK:sta seuraavin periaatein tavoitteena korpuksen mahdollisimman hyvä edustavuus ja suuri koko. Työn tavoitteiden ja siinä käytettävän vertailumenetelmän vuoksi aineisto koostuu kolmesta osakorpuksesta: 1) alkuperäissuomen korpuksesta (*AS-aineisto*), 2) monilähdekielisestä käännössuomen korpuksesta (*KSM-aineisto*) ja 3) yksilähdekielisestä korpuksista, jonka lähdekielenä on englanti kohta esitellyin perustein (*KSE-aineisto*). Koostamisperiaatteet ovat seuraavat:

a) *Kaikkien osakorpusten tekstilajina on kaunokirjallisuus*

Osakorpusten tekstilaji on kaunokirjallisuus tai tarkemmin kertomakirjallisuus, koska se ei sisällä esimerkiksi näytelmiä tai lyriikkaa (vrt. Oulun korpuksen rakenne, Saukkonen ym. 1979: 14). Korpus ei myöskään sisällä viihde- ja salapoliisikirjallisuutta; ne on Käännössuomen korpuksessa luokiteltu eri osakorpukseen. Kaunokirjallisuus on valikoitunut tutkimusaineistoksi ulkoisista lähtökohdista, eikä tutkimusintressinä ole ensisijaisesti kaunokirjallisuus sinänsä. Tekstilajin valintaan on vaikuttanut kielenulkoinen todellisuus: vain kaunokirjallisuus on sellainen tekstilaji, josta on KSK:n avulla pystytty kokoamaan kolme eri osakorpusta. Tähän vaikuttaa se tosiasia, että muuta kuin kertomakirjallisuutta käännetään harvoin muista kielistä kuin englannista, mikä vaikuttaa ratkaisevasti monilähdekielisen osakorpuksen (KSM-aineiston) koostamiseen. KSK sisältää kyllä joitakin muista kielistä kuin englannista käännettyjä akateemisia ja populaaritieteellisiä tekstejä, mutta niiden määrä on niin pieni, että työn tavoitteiden kannalta järkevän kokoisen korpuksen koostaminen käy mahdottomaksi.

b) *Kultakin tekstintuottajalta on mukana enintään yksi teos*

Jotta tutkimuksessa voitaisiin välttyä mahdollisimman hyvin yksittäisen tekstintuottajan – joko alkuperäissuomalaisen tekstin kirjoittajan tai kääntäjän – idiosynkrasian vaikutukselta, on tutkimusaineistoon otettu mukaan vain yksi teos kultakin kirjailijalta ja kääntäjältä. Lisäksi on pyritty siihen, että kunkin tekstintuottajan tekstit ovat osakorpuksittain mahdollisuuksien rajoissa suunnilleen yhtä pitkät. Tämä osoittautui kuitenkin vaikeaksi tavoitteeksi, sillä valikoinnin mahdollisuus oli vähäistä KSK:n kaunokirjallisten tekstien rajallisen määrän ja tekstien pituuden suuren vaihtelun vuoksi.

c) *Kukin korpus on suunnilleen yhtä suuri*

Korpusta koostettaessa yksittäiset osakorpuksot on pyritty saamaan mahdollisimman samansuuruisiksi. Tällä pyritään takaamaan korpusten vertailtavuus siten, että kussakin korpuksessa olisi mahdollisimman monta eri tekstintuottajalta peräisin olevaa tekstiä. Korpuksot ovat kuitenkin jossain määrin erikokoisia, joten niistä saatuja tuloksia ei voida aina vertailla puhtaiden esiintymien avulla, vaan tutkimuksessa käytetään pääasiassa suhteellisia lukuja ja tilastollista testausta.

d) *Mikään lähdekieli ei yliedustu monilähdekielisessä korpuksessa*

KSM-aineisto on pyritty mahdollisuuksien mukaan kokoamaan siten, että ensinnäkin sen lähdekielten kirjo on KSK:n rajoissa mahdollisimman suuri ja toisaalta siten, että mikään lähdekieli ei yliedustu tässä osakorpuksessa. Periaate on olennainen yleistettävyyden kannalta: koska KSM-aineiston on tarkoitus edustaa käännoosuomea yleensä, ei ole tarkoituksenmukaista, että jokin lähdekieli hallitsisi korpuksessa. Korpuksen lähdekielitasapainon vuoksi tähän korpukseen on otettu yhdestä lähdekielestä enintään kaksi teosta. Tutkimuskorpus olisi pystytty tasapainottamaan täsmällisemminkin niin lähdekielten kuin vaikkapa tekstintuottajien mukaan ottamalla korpukseen mukaan samansuuruiset tekstifragmentit. Tämä olisi kuitenkin ollut vastoin KSK:n koostamisperiaatteita, ja vielä tärkeämpi syy on se, että näin tehtäessä olisi korpuksesta tullut sanastollista tutkimusta silmällä pitäen hyvin pieni. Lähdekielikohtaisen valinnan suhteen on siis jouduttu tyytymään kompromissiin.

e) *Yksilähdekielisen korpuksen lähdekieli on englantia*

Yksilähdekielisen korpuksen valintaan vaikuttaa jälleen KSK:n rakenne ja toisaalta osittain myös suomennoskulttuuri: ainoastaan englannista suomennettujen tekstien määrä on niin suuri KSK:ssa, että voitiin luoda kahdelle muulle aineistolle saneäärältään jokseenkin verrannollinen korpus. Englantia osoittautuu hyväksi lähdekieleksi tähän aineistoon kuitenkin siksi, että angloamerikkalaista kulttuuria ja englantia voidaan pitää suhteessa suomeen ja suomalaiseen kulttuurin dominoivassa asemassa (Mauranen 2000b: 121, 2004a: 72). Tällöin Touryn (1995: 278) interferenssin toleranssia koskevaa kääntämisen lainalaisuutta soveltaen englannista suomennetuissa teksteissä siedettäisiin enemmän interferenssiä, eli käsillä olevan työn kannalta epätyypillisiä myötäesiintymiä, kuin muista aineiston lähdekielistä, jotka eivät ole yhtä hegemonisessa asemassa.³⁸

Edellisten periaatteiden mukaan tutkimuksen aineistona oleva Verrannollinen kaunokirjallisuuden korpus (VKK) muodostui seuraavanlaiseksi. AS-aineisto sisältää siis alun perin suomeksi kirjoitettuja tekstejä, ja KSM- sekä KSE-aineisto ovat käännoosuomea sisältäviä osakorpuksia. Aineisto jakautuu osakorpuksiin seuraavasti (aineiston materiaali lueteltu lähdeluettelossa):

³⁸ Interferenssin sietoa kulttuureittain on käsitellyt tarkemmin Baker (1999: 289–291). Aivan viimeaikaisessa tutkimuksessa (Mauranen 2004a) lähdekulttuurin dominanssia ei pystytty osoittamaan interferenssin määrään ja siedettävyyteen vaikuttavaksi tekijäksi.

- a) Alkuperäissuomen korpus (AS-aineisto)
 - n. 987 000 sanetta / 26 tekstiä
- b) Monilähdekielinen käännössuomen korpus (KSM-aineisto)
 - n. 1 075 000 sanetta / 13 tekstiä
 - lähdekielinä englanti, espanja, hollanti, norja, ranska, ruotsi, saksa, unkari, venäjä ja viro
- c) Yksilähdekielinen käännössuomen korpus (KSE-aineisto)
 - n. 795 000 sanetta / 11 tekstiä
 - lähdekielenä englanti

5.3. Havaintoyksiköiden valinta

5.3.1. Perusteet havaintoyksiköiden valinnalle

Edellä mainittujen hypoteesien testaamista varten on valittava joukko synonyymeiksi luettuja lekseemejä. Suomen kielen synonyymisten ilmausten suuri määrä on tutkimuksen kannalta paradoksi: valittavaksi joukoksi on tarjolla runsaasti mielenkiintoisia synonyymipesyeitä, mikä puolestaan aiheuttaa runsaudenpulan. Havaintoyksiköiden valintaa onkin lähdetty pohtimaan työn erilaisten tavoitteiden ja aineiston ominaispiirteiden avulla. Seuraavaksi esitellään ne vähimmäispiirteet, jotka havaintoyksiköiden on täytettävä sen lisäksi, että sanat ovat ylipäättään luettavissa synonyymisiksi:

- a) Valittavan synonyymipesyeen jäsenistä on tutkittaviksi päätyvien sanojen oltava tarpeeksi taajaan esiintyviä, jotta niiden kontekstuaalisia ominaisuuksia voidaan käytettävissä olevan aineiston perusteella kuvata tarpeeksi luotettavasti.
- b) Sanojen on oltava suhteellisen riippumattomia aineiston teksteissä käsitellyistä teemoista; mikäli tutkittaviksi lekseemeiksi valitaan sellaisia, joiden esiintyminen on motivoitunut käsiteltävästä aihepiiristä, on todennäköistä, että osa teksteistä ei sisällä havaintoyksiköitä ollenkaan tai taajuus jää pieneksi. Tällöin tutkittavista lekseemeistä ei saada edustavaa otosta.
- c) Lekseemien on oltava sellaisia, joilla on ainakin jonkinkaltaiset leksikaalistu-
neet vastineet käännöskorpusten lähdekielissä. Sanat eivät voi olla esimerkiksi kohdekieli- ja -kulttuuri-spesifisiä reaaleja, joille ei löydy vastinetta lähdekielistä ja -kulttuureista, sillä tällaisen sanaston edustuminen käännössuomen aineistossa osoittautuisi luonnollisesti vähäiseksi, otettakoon nyt esimerkiksi suomalaisille tuttu saunakulttuuriin kuuluva *vasta* (vrt. Tirkkonen-Conditiin hypoteesi kohdekielispesifisten ilmausten aliedustumisesta käännöskielessä, luku 3.3.1).

Edellä mainituin perustein olen päätenyt valitsemaan tutkimuksessa käytettäväksi havaintoyksiköiksi ominaisuuden astetta ilmaisevat *astemääritteet*, joista ovat käytössä myös käsitteet *intensiteettiadverbit* (Ahlman 1933; Penttilä 1963; Orpana 1988) ja toisinaan myös *ad-adjektiivit* (Hakulinen ja Karlsson 1979). Valintaan ovat vaikuttaneet myös aiemmat havaintoni astemääritteiden kontekstuaalisuudesta ja käytöstä käännöskielessä (Jantunen 1995; 2001a). On oletettavaa, että astemääritteet soveltuvat havaintoyksiköiksi hyvin. Ensinnäkin ne ovat suhteellisen taajaan esiintyviä kielenaineiksia: Saukkosen ym. (1979) taajuussanastossa esimerkiksi *hyvin* on sijaluvulla 64, *varsin* 204 ja *erittäin* 261 suomen kielen yleisimpien sanojen luettelossa³⁹. Toiseksi niiden käyttöä eivät oletettavasti ohjaa teksteissä käsiteltävät aihepiirit, joskin ne ovat tekstilaji- ja rekisterisidoksia (esimerkiksi *tosi* ja *ihan* [*tosi/ihan hyvä*] ovat tyypillisiä puhutun kielen astemääritteitä, joita tuskin tavataan esimerkiksi tieteellisistä teksteistä muuten kuin ehkä esimerkkiaineistona). Lisäksi astemääritesysteemi on olemassa sekä indoeurooppalaisissa että suomalais-ugrilaisissa kielissä, minkä vuoksi voidaan olettaa, että määritteitä tavataan myös käännösuomen aineistosta (edellyttäen, että kyseisen kategorian lähdekieliset jäsenet on käännetty vastaavilla suomalaisilla ilmauksilla).

Astemääritteet ovat tutkimuksen tavoitteiden kannalta muutenkin mielenkiintoisia. Astemääritteiden eräänlaisen ydinjoukon muodostavat leksikaalistuneet opaakki määritteet, jollaisiksi Tuomikoski (1973: 208) mainitsee muun muassa määritteet *hyvin*, *kovin* ja *oikein*. Tällaiset määritteet sijoittuvat leksikaalisten ja kieliopillisten sanojen rajamaastoon, sillä niillä on sekä sisältö- että kieliopillisten sanojen piirteitä. Samaan ominaisuuteen on kiinnitetty huomiota myös englannin kielen astemääritteiden tutkimuksessa: esimerkiksi englannin *very*-astemääritettä pidetään joissakin tutkimuksissa enemmänkin kieliopilliseen sanastoon kuin sisältösanoihin kuuluvana (ks. mm. Bolinger 1972: 18; Bäcklund 1973: 7). Leksikaalistuneita opaakkeja astemääritteitä voinee pitää jollei kieliopillisina niin ainakin puolikieliopillisina sanoina (vrt. Partington 1998: 32). Synonymian tutkimisen kannalta tämä on mielenkiintoista muun muassa siksi, että kieliopilliset sanat eivät voisi muodostaa synonymiasuhteita (Kuiiri 2000; ks. aiemmin lukua 4.1).

Mikäli leksikaalistuneita astemääritteitä pidetään kieliopillistuneina, ne ovat selvä osoitus kieliopillistuneiden sanojen sisäisestä synonymiasta. Kieliopillisia sanoja astemääritteet lähestyvät siinä, että ne suhtautuvat muotosanojen tavoin kielenulkoiseen maailmaan: astemääritteillä ei ole viittaussuhdetta kielenulkoiseen tarkoitteeseen. Lisäksi ne eivät voi esiintyä lauseessa itsenäisenä jäsenenä vaan ovat aina sidoksissa (määrityssuhteessa) johonkin toiseen lekseemiin. Toisaalta ryhmän avoimuus eli mahdollisuus saada uusia jäseniä on samankaltainen piirre sisältösanojen kanssa (Klein 1998: 27–28). Leksikaalistumista tapahtuu adjektiivien *n*-muotojen (esim. *valtava-n*) kvalitatiivisen merkityksen hävitessä, jolloin määritteet leksikaalistuvat pelkästään kvantitatiiviseen määrityssuhteeseen. Näin on käynyt esimerkiksi sellaisille määritteille kuin *hirveän* ja *tavattoman*, joiden yhdistymistä eivät enää ohjaa määritteiden kvalitatiiviset ominaisuudet. Niinpä ne voivatkin esiintyä sellaisissa määrityssuhteissa, joissa määritteen ja määritettävän välillä olisi

³⁹ Tutkimuksen tekohetkellä ei ole ollut saatavilla tuoreempaa, esimerkiksi Suomen kielen tekstipankkiin perustuvaa sanojen yleisyyslistaa.

havaittavissa semanttinen ristiriita, elleivät määritteet olisi leksikaalistuneet, esimerkiksi *hirveän kaunis, tavattoman pieni*. (Orpana 1988: 166–170.)

Kontekstuaalisten valintojen tutkimisen kannalta on kiinnostavaa, että leksikaalistuneiden ja erityisesti tavallisimpien astemääritteiden liittymisvapaus on suuri, sillä niitä eivät rajoita ilmausten kvalitatiiviset ominaisuudet, kuten on laita monilla transparenteilla *n*-muotoisilla määritteillä, joissa kvalitatiivinen merkitys on vielä tallella (*?juhlallisen epätavallinen nainen, ?tappavan lyhyet hiukset*). Opaakit astemääritteet ovat sen sijaan kadottaneet kvalitatiiviset ominaisuutensa, ja niiden funktiona on pelkästään ominaisuuden asteen ilmaiseminen. Tällaisten määritteiden kombinaatiot eivät ole kvalitatiivisten piirteiden säätelemiä, vaan neutraaleina asteen ilmaisijoina niillä on suuri liikkumavapaus (ks. Orpana 1988: 166–). Tämä tarkoittaa myös sitä, että niiden kontekstuaaliset ominaisuudet eivät ole kovinkaan helposti selvitettävissä pelkästään intuitioon nojautumalla. Jos Smadjan (1993: 147) lausuma, jonka mukaan kollokaation toisen jäsenen voi ennustaa, kun toinen jäsen on tiedossa, pitää paikkansa, olisivat astemääritteet tässäkin suhteessa oivallisia tutkimuskohteita, sillä yleisimpien määritteiden kollokationaalista valintaa ei voine tarkasti tietää sisäänrakennetun kielitiedon perusteella. Toisin on kuitenkin tilanne esimerkiksi *tuiki-* ja *ratki-*määritteillä, joista ensin mainitulle intuitio valinnee kollokaatiksi luultavasti hyvin yleisesti *tavallisen* ja jälkimmäiselle *riemukkaan*. Myös frekventille *hyvin-*määritteelle voisi lueta moniakkin kollokaattikandidaatteja, mutta mahdollisuuksien valikoima osoittautunee kuitenkin niin suureksi, ettei intuitio enää riitä kollokationaalisten valintapreferenssien löytämiseen. Lisäksi on muistettava, että kontekstuaalinen valinta sisältää joukon muitakin ominaisuuksia kuin leksikaaliset yhteisesiintymät, jotka lienevät helpoimmin mieleen tulevia.

5.3.2. Astemääritteet

Astemääritteet ovat etumääritteitä, jotka määrittävät pääasiassa adjektiiveja (*melko ikävä*) ja adverbbeja (*kamalan nopeasti*) (Vilkuna 1996: 207–209). Ne voivat määrittää myös kvanttoreita, kuten Hakulisen ja Karlssonin (1979: 83) esimerkissä *Löy-simme erittäin paljon mustikoita*. Astemääritteiden käyttöä määriteltäessä ei yleensä tule esiin niiden mahdollisuus määrittää myös adpositioita (pre- ja postpositioita). Tällainen määrätyssuhde on nähtävissä kuitenkin Hakulisen ja Karlssonin adpositioihin keskittyvässä analyysissä, jossa he pohtivat, määrittääkö astemäärite koko prepositiorakennetta [*aivan [ilman apua]*] vai pelkästään prepositiota [[*hyvin lähellä*] *kuolemaa*]. Esimerkkitapaukset paljastavat, että astemääritteet voivat toimia myös preposition tai prepositiolausekkeen määritteenä. Postpositiot eivät voisi kuitenkaan saada määritteitä. (Mts. 154–155.) Tämä pitääkin paikkansa, sillä rakenteet kuten **selän aivan takana* ovat kielenvastaisia. On kuitenkin mahdollista, että astemäärite määrittää koko postposition sisältämää lauseketta: *Hän seisoi aivan selkäni takana*. Tässä ei määrätyssuhteeksi voi tulkita rakennetta [[*aivan selkäni*] *takana*] vaan se on jäsennettävä toisin: [*aivan [selkäni takana]*].

Astemääritteiden kontekstianalyyseja on tehty suomesta niukalti, vaikka edellä kerrottu kävisikin osittain kolligaattianalyysin pohjaksi, sillä astemääritteiden oikeanpuoleisiksi vieruskolligaateiksi voi sen perusteella olettaa adjektiivit, adverbit, kvantorit ja adpositio(lausekkeet). Varsinaisia luonnolliseen aineistoon perustuvia astemääritteisiin keskittyviä kontekstianalyyseja on tehty jonkin verran englannin kielestä, joskin ne ovat lähes yksinomaan astemääritteen ja niiden pääsanaposition lekseemien muodostamien kollokaatioiden analyysia (ks. mm. Bolinger 1972; Bäcklund 1973; Altenberg 1991; Johansson 1993; Paradis 1997). Bäcklundin (1973) englannin kielen astemääritteitä tarkastelevasta tutkimuksesta on myös löydettävissä viitteitä kolligaatioihin liittyvistä havainnoista, vaikka tutkimus ei pyrikään tarkastelemaan astemääritteiden kolligaatiosuhteita. Tutkimuksesta (mts. 282) voi esimerkiksi lukea, että lähisyronyymeilla *completely* ja *fully* ('täysin, kokonaan') on adjektiivipääsanojensa joukossa eri määrä verbin partisiippimuotoja: *completely*-määritteellä niitä on 35 % kun taas *fully*-määritteellä 62 %.

Määritteiden ja pääsanojen semanttisia kytkentöjä on suomesta tarkastellut Orpana (1988) ja pääasiassa hollannista ja englannista Klein (1998). Nämä eivät kuitenkaan ole varsinaisia kontekstianalyyseja, vaan ne tarkastelevat määrittysuhteiden mahdollisuutta ja luonnetta määritteen ja pääsanana semanttisten ominaisuuksien pohjalta. Kontekstuaalisia (lähinnä kollokaatteihin ja semanttiseen preferenssiin liittyviä) piirteitä voi näistäkin analyyseista poimia, mutta tutkimukset eivät kuitenkaan pyri tyhjentäviin kontekstia kuvaaviin esityksiin. Mäkipään (1995) analyyseista on vuorostaan löydettävissä piirteitä, jotka lähestyvät morfologisen preferenssin analyyseja, sillä tutkimuksessa määritteiden ja pääsanojen yhdistymismahdollisuuksia tarkastellaan osittain pääsanana olevien adjektiivien vertailuasteiden perusteella: positiivimuotoisilla adjektiiveilla on omanlaisensa määritteet (esim. *hyvin* ja *sangen*) samoin kuin komparatiivi- (*paljon, yhä*) ja superlatiivimuotoisillakin (*kaikkein*).

Astemääritteiden ryhmästä olen valinnut havaintoyksiköiksi niin sanotut vahventajat eli määritteet, jotka ilmoittavat jonkin ominaisuuden hyvin suurta astetta. Tutkimuskohteeksi ne sopivat sikäli hyvin, että ryhmä sisältää paljon synonymiaa, mikä näkyy muun muassa Leinojen *Synonyymisanastosta* (Leino ja Leino 1990) ja Orpanan (1988) sekä Mäkipään (1995) tutkimuksista. Ilmauksen astetta vahventavia määritteitä ovat muun muassa *hyvin, kovin, hirveän, tavattoman* ja *jumalattoman*. Näistä ainoastaan kaksi ensimmäistä ovat opaakkeja, muiden määritteiden sisältö on emotionaalinen (Orpana mts. 167). Olen ottanut lähtökohdaksi *hyvin*-määritteen, joka on yleisin vahvistaja Saukkosen ym. (1979) taajuussanakirjan mukaan ja jota myös Vilppula (2001) epäilee yhdeksi kaikkein yleisimmistä vahvistajista.

hyvin-määritteen synonyymeiksi esitettyjä määritteitä olen etsinyt kahdesta yleiskielen sanakirjasta ja Leinojen *Synonyymisanastosta*. Ensinnäkin *Nykysuomen sanakirjan* sana-artikkelista saadaan seuraavat tiedot:

hyvin adv.

2. vahvistavana, tehostavana

b. perusasteessa olevaa adjektiivia ja adverbia sekä partisiippiä määrittäen: *erittäin, sangen, kovin, varsin* | *H. suuri, korkea, kirkas, hyvä, arvokas, harvinainen. H. pehmeää puuta. H. varhainen aamu. H. tavallinen asia. On h. mahdollista, että --. H.*

pieni, ahdas, huono, vaikea. Hän tuli h. onnettomaksi. Se on h. epätietoista. Keskusteltiin h:kin erilaisista asioista. Onko se h:kin kallis? – H. edustava seurue. H. tunnettu liike. On h. luultavaa, toivottavaa, että --. Se on mielestäni h. ymmärrettävää. – H. nopeasti, huonosti. H. kaukana, myöhään. Loukkaantui h. vaarallisesti. Hän oli h. mielissään. H. vähän.
 c. harv. substantiivina määrittäen: *Hevosien on h. jano 'kova jano'. Se tapahtui h. keväällä 'varhain keväällä'. Lahden pohjukassa oli jo h. hämy.*

NS:n sana-artikkelista on havaittavissa, että *hyvin* voi ensinnäkin määrittää monenlaisia adjektiiveja, kuten leksikaalistuneelle määritteelle voi olettaakin olevan mahdollista. Lisäksi esimerkeistä on löydettävissä määrittämissuhteita, joissa pääsanoina on adverbeja ja kvanttoireita. Harvinaiset substantiivin määrittäystä osoittavat tapaukset lienevät selitettävissä *jano-*, *keväällä-* ja *hämy-*substantiivien adjektiivimaisuudella ja adverbimaisuudella näissä tapauksissa, ja harvinaisina ne lienevät enemmänkin luovan kielenkäytön tuotoksia. Määritteen synonyymeiksi on NS:ssä mainittu *erittäin*, *kovin*, *sangen* ja *varsin*.

Suomen kielen perussanakirja (CD-PS 1997) määrittelee *hyvin*-lekseemin lyhyemmin, mutta noudattaa suurelta osin NS:n määrittelyä:

hyvin

2. vahvistus- ja tehostussanana.

b. erittäin, kovin, varsin, sangen. Hyvin suuri, harvinainen, huono. Hyvin moni heistä. Hyvin nopeasti. Hyvin paljon, vähän. On hyvin luultavaa, että --. Onko se hyvinkin kallis?

Leinojen *Synonyymisanasto* (1990) ei taas anna määritteelle käyttöesimerkkejä vaan tyytyy luettelemaan määritteen merkitystä selittäviä synonyymeja, joita onkin annettu kosolti:

hyvin

aivan, erittäin, ihan, juuri, kerrassaan, kokonaan, kovin, lopen, oikein, perin, sangen, suorastaan, tarkalleen, tykkänään, tyystin, tyyten, täysin, vallan, varsin, ylen.

Kolmen sanakirjan perusteella tavallisimmiksi synonyymeiksi erottuvat selvästi *erittäin*, *kovin*, *sangen* ja *varsin*. Leinojen synonyymiluettelosta on mainittava, että se sisältää runsaasti myös maksimoijia eli ominaisuuden suurinta astetta ilmoittavia astemääritteitä (*aivan, ihan, juuri, kerrassaan, kokonaan, tykkänään, tyystin, tyyten, täysin, vallan*), joita ei mielestäni voi suoranaisesti laskea *hyvin*-määritteen synonyymeiksi, joskin rajat maksimaalisen ja vahventavan asteen ilmoittamisen välillä voivat olla häilyviä (kyseisentyypisiä määritteitä pitää myös Mäkipää [1995: 95–97] ominaisuuden suurinta astetta ilmaisevina). Mainituille neljälle määritteelle löytyy edelleen synonyymisia opakkeja leksikaalistuneita astemääritteitä kyseisten sanojen sana-artikkeleista. Näitä ovat edellisistä sanastoista yhteen koottuna *järin*, *lopen*, *oikein*, *perin* ja *ylen*. Sanakirjoista ei synonyymien listalta löydy kuitenkaan astemääritteitä *todella* ja *tosi*, jotka ovat samantapaisia intensiteettisanoja ja joista etenkin jälkimmäinen on suosittu nykypuhekielessä vahvistajana käytettynä (*tosi-*

sanasta mm. Punttila 1990: 19; Lappalainen 2000: 80). *todella*-adverbia käytetäänkin CD-PS:n (s.v. *todella*) mukaan usein vahvistusmääritteen luonteisena (käyttöesimerkkeinä *Ruoka oli todella hyvää. Hyppäsi todella pitkälle*), ja vahventajaksi sen lukee myös 1990-luvun aikakauslehtikieltä tarkastellut Mäkipää (1995: 93). CD-PS:ssa on *tosi*-määritteen synonyymeiksi lueteltu määritteet *todella* ja *erittäin* (s.v. *tosi*; käyttöesimerkkeinä *Tosi hyvä, huono, vanha, arvokas. Tosi nopeasti. Sain sen tosi halvalla.*), joten sitäkin on pidettävä samaan synonyymipesyeeseen kuuluvana vahventajana. Koska tutkimusmateriaali koostuu suurimmaksi osaksi 90-luvun jälkipuoliskolla tuotetuista teksteistä eli nykykielestä, on perusteltua sisällyttää nämäkin astemääritteet seuraavaan valintavaiheeseen, jossa lopulliset tutkimuksessa käytettävät havaintoyksiköt määräytyvät. Tähänastisessa valintaprosessissa on vahventajien joukoksi muodostunut seuraavien astemääritteiden muodostama ryhmä: *erittäin, hyvin, järin, kovin, lopen, oikein, perin, sangen, todella, tosi, varsin* ja *ylen*.

Varsinaisten havaintoyksiköiden valitsemiseksi olen hakenut edellä mainitut leksikaalistuneet astemääritteet tutkimusaineiston alkuperäissuomen osakorpuksesta (AS-aineisto) tarkoitukseni valita niistä sellaiset, jotka ovat tarpeeksi taajaan esiintyviä tutkimusmateriaaliksi. Seuraavaan taulukkoon on koottu astemääritteiden frekvenssit AS-aineistossa:

Taulukko 5.1. Leksikaalistuneiden vahventajien taajuus AS-aineistossa.

vahventaja	f
<i>hyvin</i>	352
<i>kovin</i>	176
<i>oikein</i>	121
<i>todella</i>	82
<i>varsin</i>	64
<i>erittäin</i>	50
<i>tosi</i>	44
<i>sangen</i>	9
<i>perin</i>	6
<i>järin</i>	5
<i>ylen</i>	4
<i>lopen</i>	1

Vahventajien frekvenssit ovat aineistossa hyvin erilaisia: ne vaihtelevat *lopen*-määritteen yhdestä esiintymästä *hyvin*-määritteen runsaaseen kolmeen ja puoleen sataan esiintymään. Luetelluista määritteistä valitsen havaintoyksiköiksi kolme frekventeintä eli tummennetut *hyvin*-, *kovin*- ja *oikein*-määritteet, koska tutkittavien yksiköiden on oltava tarpeeksi taajaan esiintyviä, jotta niiden kontekstuaalisista ominaisuuksista pystytään tekemään päteviä päätelmiä. Kolmen synonyymien vertailu antaa kuvauskohteesta tarkemman kuvan kuin vain kahden havaintoyksikön tarkastelu. Toisaalta muiden kuin valittujen synonyymisten määritteiden frekvenssi on niin pieni, että niiden kontekstuaalisten valintapiirteiden tarkastelu tuottaisi ongelmia. Tämä todellistuisi erityisesti leksikaalisten myötäesiintymien tarkastelussa, mikä vaatii pohjakseen yhtäältä laajan aineiston ja toisaalta suuren määrän esiintymiä.

5.4. Tutkimusmenetelmät

5.4.1. Vertailumenetelmät

Tutkimukseni raportointiosan ensimmäinen osa (luku 6) eli astemääritteiden synonyymisyyttä kontekstuaalisesta näkökulmasta käsittelevä osuus perustuu kokonaisuudessaan alkuperäissuomen osakorpukseen eli AS-aineistoon. Synonyymisyyttä ja kontekstuaalisia valintoja tutkitaan jaksossa siten, että astemääritteitä vertaillaan keskenään ja pareittain. Tarkoituksena on luoda kokonaiskuva siitä, miten astemääritteet mahdollisesti poikkeavat kontekstuaalisilta muuttujiltaan toisistaan. Lisäksi pareittain tarkastelu antaa mahdollisuuden katsoa, mitkä näistä kolmesta määritteestä ovat keskenään synonyymisimmät ja mitkä taas poikkeavat kontekstuaalisilta piirteiltään eniten toisistaan.

Käännössuomen erityispiirteitä käsittelevä luku 7 perustuu taas tutkimusaineiston kaikkien osakorpusten vertailuun. Vertailun tarkoituksena on tutkia, a) poikkeaa ko käännöskieli yleisesti alkuperäiskielestä ja b) miten yksi lähdekieli mahdollisesti vaikuttaa kohdekieleen. Käytetyssä analyysissä on heijastumia aiemmista vertailumenetelmistä, joilla on pyritty analysoimaan käännöskielen erilaisuutta suhteessa alkuperäiskieleen ja myös yksittäisen lähdekielen vaikutusta. Olen luetellut tutkimuksissa hyödynnettyjä, erityisesti lähdekielen vaikutuksen selvittämiseen käytettyjä menetelmiä taulukkoon 5.2.

Taulukko 5.2. Tutkimusmenetelmiä, joilla on pyritty tarkastelemaan lähdekielen vaikutusta.

Tutkimus	Menetelmä
Laviosa (Laviosa-Braithwaite 1996, Laviosa 1998)	a) Kieliryhmäpohjaisten lähdekielisten korpusten vertailu. b) Kaksi- ja yksilähdekielisten korpusten vertailu.
Mauranen (2000b) ⁴⁰	Alkuperäiskielisen korpuksen vertailu a) yksilähdekieliseen (englanti) korpukseen b) monilähdekieliseen korpukseen (ei englantia).
Eskola (2002)	Kahden yksilähdekielisen korpuksen vertailu (englanti ja venäjä).

Kuten taulukko kertoo, lähdekielen vaikutuksen analysointi ei ole korpuslingvistikissä menetelmissä tavoitteena uutta ja tutkimusmenetelmiä on useita. Laviosan (Laviosa-Braithwaite 1996, Laviosa 1998) ensisijainen tarkoitus on ollut tarkastella nimenomaan käännösendlannin eroa alkuperäisenglantiin verrattuna, mutta hän on pyrkinyt ottamaan huomioon myös lähdekielen vaikutuksen. Tämän saavuttamiseksi Laviosa on verrannut toisaalta kieliryhmäpohjaisia lähdekielisiä korpuksia keskenään (esim. germaanisista ja romaanisista kielistä käännettyjä tekstejä) ja toisaalta kaksilähdekielisiä korpuksia (esim. italiasta ja espanjasta käännettyjä teks-

⁴⁰ Myöhemmin Mauranen (2004a) on lisännyt vertailuun myös venäjistä suomennettujen tekstien yksilähdekielisen korpuksen.

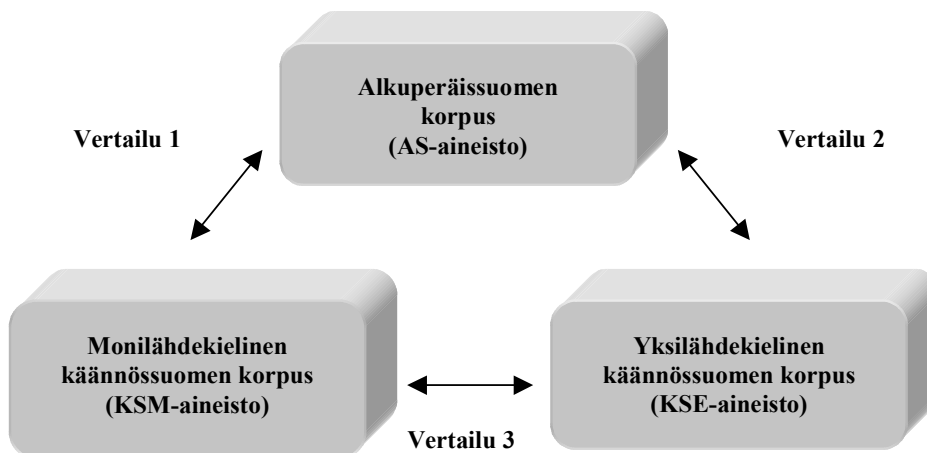
tejä) yksilähdekielisten korpusten (mm. ranskasta käännettyjen tekstien) kanssa. Menetelmä, kuten on huomattavissa, ei kuitenkaan pysty kertomaan erityisesti *yhden* lähdekielen vaikutusta kohdekieleen, koska vertailtavana olevat korpukset sisältävät useita lähdekieliä – niin sukulaiskieliä kuin ovatkin. Menetelmän arvo on kuitenkin nähtävä ensimmäisenä askeleena kohti verrannollisten korpusten ja tutkimusmenetelmien kehittämistä.

Maurasen (2000b, ks. myös 2004a) menetelmä tavoittaa paremmin nimenomaan yhden lähdekielen vaikutuksen tutkimisen, sillä hänellä on vertailussa mukana yksilähdekielinen korpus (suomennoksia englannista), jota tarkastellaan suhteessa alkuperäiskieliseen korpukseen. Tämän vertailun tuloksia verrataan vuorostaan monilähdekielisen ja yksilähdekielisen korpusten vertailusta saatuihin tuloksiin. Mahdolliset erot indikoivat lähdekielen vaikutusta, yhtäläiset tulokset taas päinvas- taista tilannetta. Viimeisin tässä esiin nostettava menetelmä on Eskolan (2002) metodi, jossa hän vertaa kahden yksilähdekielisen aineiston tuottamia rakenteita ja frekvenssejä. Eskolan menetelmä poikkeaa ratkaisevasti edellisistä (ja myös tässä tutkimuksessa käytetyistä), koska siinä ei ole käytetty monilähdekielistä korpusta, eli käännöskielen olemusta ei ole pyritty lähestymään niin sanotun yleisen käännös- suomen kautta vaan induktiivisesti kahdesta lähdekielestä olevan aineiston avulla.

Oman tutkimukseni vertailumenetelmä pohjautuu edellä mainittuihin Laviosan ja Maurasen käyttämiin menetelmiin. Näille molemmille on yhteistä yhtäältä monilähdekielisen korpusten käyttäminen ja toisaalta lähdekielen vaikutuksen tarkas- telu (ainakin jossain määrin) lähdekielispesifisen korpusten avulla. Menetelmäni on lähellä erityisesti Maurasen käyttämää, joskin pienin muutoksin: ensinnäkin moni- lähdekielinen korpus sisältää myös englannin lähdekielenä, koska tämän korpusten on tarkoitus edustaa suomennoksia yleensä, ei vain suomennoksia, joita on tehty muista kielistä kuin englannista. Monilähdekielisen korpusten on mielestäni sisäl- lettävä kuitenkin myös englannista suomennetut tekstit. Tämä on olennaista muun muassa siksi, että englannista suomennettujen tekstien määrä on ylivoimaisesti suu- rin ryhmä suomennosten joukossa (ks. Minkkinen 2001: 8), joten englannin puuttu- essa KSM-aineisto ei edustaisi suomalaista käännöstodellisuutta. Tämän aineiston on tarkoitus antaa mahdollisimman monipuolinen kuva käännössuomesta, ja eng- lannin pois jättäminen kaventaisi lähdekielten kirjoa. Lisäksi KSM-aineisto ilman englantia lähdekielenä muuttaisi vertailun sellaiseksi, jossa verrataan englannista käännettyjä tekstejä muista kielistä käännettyihin teksteihin, mikä ei ole kuitenkaan tutkimuksen tarkoitus.

Omaan menetelmäni olen lisännyt monilähdekielisen ja alkuperäissuomen kor- pusten sekä yksilähdekielisen ja alkuperäissuomen korpusten vertailun lisäksi myös kolmannen vertailun, nimittäin käännössuomen korpusten vertailun (vertailu 3 kuviossa 5.1). Koska vertailu 1 keskittyy alkuperäissuomen ja käännössuomen tarkasteluun ja vertailu 2 puolestaan alkuperäissuomen vertailuun suhteessa englan- nista käännettyihin teksteihin, jäisi englannista suomennettujen ja yleensä käännös- kielen vertailu implisiittiseksi ilman kolmannen vaiheen mukaan ottamista. Tutki- muksessa ei siis keskitytä tarkastelemaan pelkästään käännöskielen suhdetta alku- peräiskieleen, vaan siinä otetaan huomioon myös käännöskieliaineistojen keskinäi- nen vertailu. Näin ollen lähdekielen vaikutusta testataan analyysissä kahdella ta-

valla: vertaamalla vaiheiden 1 ja 2 tuloksia toisiinsa sekä vertaamalla kahta käännöskorpusta keskenään.



Kuvio 5.1. Verrannollinen kaunokirjallisuuden korpus ja kolmivaiheinen vertailumenetelmä.

5.4.2. Tilastolliset menetelmät

Tutkimukseni on luonteeltaan hyvin kvantitatiivinen, kuten korpustutkimus yleensäkin. Astemääritteiden synonyymisyyttä ja kielimuotoikohtaista vaihtelua tarkastellaan seuraavissa luvuissa pitkälti punnitsemalla eri muuttujien frekvenssien ja osuuksien merkitystä ja osuuksien erojen tilastollista merkitsevyyttä. Tarkastelussa on mukana kuitenkin vahvasti myös kvalitatiivinen ulottuvuus. Kretschmarin, Meyerin ja Ingegnerin mukaan tilastollinen testaus ei saisi olla tutkimuksessa itseisarvo vaan pikemminkin pohdinnan lähtökohta:

A statistically significant result never explains the quantitative distribution under study – the analyst is always responsible for explanations – but just suggests to the analyst which distributions are probably worthy of explanations (Kretschmar ym. 1997: 175).

En siis tyydy pelkästään esittämään tarkastelussa esiin tulevia suhde- ja frekvenssieroja, vaan pyrin löytämään myös ne laadulliset tekijät, jotka ovat kvantitatiivisten erojen takana. Yksistään kvantitatiivisiin tuloksiin keskittyvä analyysi ei mielestäni pysty toteuttamaan työn tarkoitusta, sillä kuva jäisi silloin yksipuoliseksi aineiston jäsenyykseksi ja kvantitatiivisten tulosten taustalla olevat syyt jäisivät

selvittämättä. Kvantitatiivinen analyysi perustuu tilastolliseen testaukseen, jota kuvailen seuraavaksi.

Koska tarkoituksenani ei ole kuvata satunnaisia leksikaalisia kombinaatioita vaan keskittyä myötäesiintymiin, jotka ovat toistuvia ja siten tyypillisiä kullekin havaintoyksikölle ja kielimuodolle, olen päättänyt jättämään huomiotta harvoin esiintyvät kollokaatit. Koska tavoitteenani on tarkastella toistuvia myötäesiintymiä ja koska myöhemmin esitetyin perustein analyysissa käytettävä *MI*-testi suosii herkästi harvinaisia, jopa yksittäisiä, myötäesiintymiä (Clear 1993: 280), olen asettanut yhteisesiintymien vähimmäistaajuudeksi viisi esiintymää; toisaalta korpuksen suhteellinen pienuus leksikaaliseen analyysiin ei anna mahdollisuutta nostaa vähimmäisfrekvenssiä juurikaan tätä korkeammaksi.⁴¹ Vähimmäistaajuutta ei ole suhteutettu osakorpuksen kokoon, sillä aineistot ovat suunnilleen samankokoisia. Koska tutkimuksen tarkoituksena ei ole myöskään esitellä yksittäisten tekstintuottajien idiosynkraattisia valintoja, olen poiminut varsinaiseen kollokaatioanalyysiin vain sellaiset yhteisesiintymät, jotka esiintyvät vähintään kahdessa eri tekstissä eli vähintään kahden eri kirjailijan tai kääntäjän tuotannossa. Tämän jälkeen tarkasteluun seuloutuneet kollokaatit on alistettu tilastolliseen testaukseen.

Eräänlaisen tilastollisen lähestymistavan esittää Sinclair (1987, ks. myös 1991) jaotellessaan kollokaatiotapaukset kahteen ryhmään, ylös- ja alaspäin suuntautuviin kollokaatteihin. Ylöspäin suuntautuville kollokaatioilla Sinclair (1987: 325–326) tarkoittaa sellaisia, joissa kollokaatin frekvenssi on korpuksessa suurempi kuin hakusanan frekvenssi, vastaavasti alaspäin suuntautuvat kollokaatiot ovat sellaisia, joissa kollokaatin frekvenssi on korpuksessa pienempi kuin noodin frekvenssi. Sinclairin mukaan ylöspäin suuntautuvat kollokaatit ovat näistä kahdesta tilastollisesti heikompia suhteita, koska on todennäköistä, että suurifrekvenssiset kollokaatit esiintyvät todennäköisemmin noodin yhteydessä. Tällaisina kollokaatioina Sinclair pitää esimerkiksi noodien ja kieliopillisten sanojen kombinaatioita. Alaspäin suuntautuvat kollokaatiot sen sijaan kuvaavat hänen mukaansa noodin ja kollokaatin semanttisia suhteita.⁴² Ylös- ja alaspäin suuntautuvien kollokaattien tilastoimiseksi Sinclair (mp.) esittää laskentakaavaa, jossa marginaaliksi on otettu 15 %. Sen mukaan 1150 kertaa tai useammin tavattavat kollokaatit ovat ylöspäin suuntautuvia ja 850 kertaa tai harvemmin esiintyvät alaspäin suuntautuvia kollokaatteja, jos noodi esiintyy 1000 kertaa. Väliin jääviä Sinclair pitää neutraaleina. Testatakseni Sinclairin mallia olen laskenut mukaan valitsemistani astemääritteistä kahden harvinaisimman eli *kovin-* ja *oikein-*määritteiden ylös- ja alaspäin suuntautuvat kollokaatit vähintään viisi kertaa tarkastelualueella esiintyvien kollokaattien joukosta. Laskentamallin mukaan raja-arvoiksi muodostuvat taulukon 5.3 mukaiset luvut.

Raja-arvojen mukaan laskettu tulos on jokseenkin yllättävä: kaikki näiden kahden astemääritteen vähimmäisesiintymistaajuuden ylittävistä kollokaateista listautuvat ylöspäin suuntautuviksi kollokaateiksi eikä siten yksikään alaspäin suuntautuvaksi. Tulos on kuitenkin ainakin jossain määrin odotettavissa, koska kyseisten

⁴¹ Yhteisesiintymän vähimmäistaajuus vaihtelee melkoisesti tutkijoittain: Kjellmerillä (1984, 1987) se on kaksi, Stubbsilla (1995a) kolme, Kennedyllä (1991) neljä ja Jonesilla ja Sinclairilla (1974) peräti kymmenen esiintymää.

⁴² Clear (1993:281–282) pitää oman aineistonsa valossa Sinclairin tulkintaa yksinkertaistettuna.

astemääritteiden frekventit kollokaatit ovat suurelta osin kieliopillisia sanoja (ks. taulukkoa 6.5 s. 91), ja tällaiset olisivat Sinclairin mukaan nimenomaan ylöspäin suuntautuvia. Sinclairin esittämä laskentamalli osoittautuu kuitenkin liian karkeaksi tähän tutkimukseen, koska sen avulla ei pystytä tekemään kollokationaalista eroa tutkimukseen valittujen synonyymien välille, mikä taas on yksi tämän työn tavoitteista. Niinpä laskentamalli on hylättävä soveltumattomana.

Taulukko 5.3. Ylös- ja alaspäin suuntautuvien kollokaatioiden raja-arvot, esiintymätaajuudet taulukosta 5.1.

	raja-arvot	
	ylöspäin suuntautuvat	alaspäin suuntautuvat
<i>kovin</i>	202 = 176+26 (15 %)	150 = 176-26 (15%)
<i>oikein</i>	139 = 121+18 (15 %)	103 = 121-18 (15 %)

Tarjolla olevien laskentamallien määrä on korpuslingvistiikan kehittymisen ja tilastollisten menetelmien soveltamisen myötä kasvanut suureksi (erilaisista käytettävissä olevista menetelmistä ks. esim. Oakes 1998: 162–192; Kilgarriff 2001). Barnbrook (1998) esittelee kolme tilastollista menetelmää, joita on käytetty kollokaatiosuhteiden tilastollisen merkitsevyyden mittaamisessa useimmin. Menetelmät ovat *z*-, *t*- ja *MI* (*Mutual Information*) -testit⁴³. Näistä *z*- ja *t*-testit perustuvat todellisten ja odotuksenmukaisten frekvenssien vertailuun keskihajonnan pohjalta. Testien suuri testisuure kertoo, miten varmoja voidaan olla, että tarkastelualueella havaittu kollokaatti esiintyy alueella tiheämmin kuin koko aineistosta. Testit mittaavat siten varmuutta, jolla voidaan sanoa sanojen välillä olevan sattumaa suurempi myötäesiintyminen. Alhainen suure puolestaan ilmaisee, että kollokaatti ei esiinny tarkastelualueella sen tiheämmin kuin aineistossa yleensäkin. *MI*-testi taas vertailee noodin ja kollokaatin yhteisesiintymän todennäköisyyttä todennäköisyyteen, että ne olisivat toisistaan riippumattomia. Testin avulla voidaan kertoa, miten todennäköistä on, että toisen sanan esiintyminen tekee todennäköiseksi myös jonkin toisen sanan esiintymisen lähikotekstissa. Testisuure ilmaisee siis yhteisesiintymän voimakkuuden, todennäköisyyden esiintyä yhdessä, mutta ei kerro kollokaatin sattumaa suuremmasta esiintymisestä tarkastelualueella. (Barnbrook mts. 95–99; Clear 1993: 278–282.)

Vertaillen näitä kolmea testiä Barnbrook havaitsee, että *z*- ja *MI*-testit antavat suhteellisen samankaltaiset tulokset, joista *t*-testi vuorostaan poikkeaa. Tarkastelussa tulee ilmi, että kaksi ensin mainittua nostavat herkemmin merkitseviksi sellaiset kollokaatit, joiden jäsenet ovat korpuksessa harvinaisia mutta esiintyvät kuitenkin taajaan yhdessä, *t*-testi puolestaan nostaa merkitseviksi sellaiset kollokaatiosuhteet, joissa (ainakin) kollokaatti on korpuksessa frekventti. Niinpä *t*-testi poimiikin herkästi merkitseviksi kieliopilliset sanat, jotka ovat aineistoissa taajaan

⁴³ Testien kaavat ovat: $z = (O-E)/\sigma$, $t = (O-E)/\sqrt{O}$, $I = \log_2 (O/E)$, joissa *O* = lekseemin havaittu frekvenssi, *E* = lekseemin odotettu frekvenssi ja σ = esiintymien keskihajonta koko aineistossa (Barnbrook 1996: 95–99).

esiintyviä. *MI*- ja *z*-testit puolestaan painottavat herkemmin leksikaalisten sanojen suhteita. (Clear 1993: 280–281; Barnbrook 1998: 100–101; Stubbs 1995a: 39.)

Testeillä on myös ongelmansa. Sekä *z*- että *MI*-testi korostavat tapauksia, jotka ovat jopa erittäin harvinaisia ja voivat johtua idiosynkrasiasta. Ne myös jättävät huomiotta usein toistuvat myötäesiintymät, jotka kuitenkin ovat kielen kuvauksen kannalta tärkeitä, koska ne ovat taajaan käytettyjä. *t*-testi puolestaan painottaa suuri-frekvenssisia tapauksia ja jättää huomiotta harvoin esiintyvät, mutta ehkä mielenkiintoiset kollokaatiosuhteet. (Barnbrook 1998: 101; Clear 1993: 280–281.) Sekä Barnbrook (mp.) että myös Stubbs (1995a: 40) ehdottavat useamman kuin yhden testin käyttöä, koska mittausmenetelmät kuvaavat osin eri asioita. Tällöin pystytään myös vähentämään yhden testin aiheuttamaa vääristymää. Valitsenkin työni kollokaatioanalyysiin yhden testin sijasta kaksi erilaista testiä, sekä *t*- että *MI*-testin, jotka poikkeavat toisistaan sen perusteella, mitä ne kertovat yhteisesiintymästä ja miten ne ottavat huomioon esiintymistiheyden. Tutkimuksessa on lisäksi pyritty häivyttämään *MI*-testille ominaista harvinaisten ja mahdollisesti idiosynkraattisten tapausten painottumista pitämällä vähimmäisesiintymistäajuus suhteellisen korkealla ja valitsemalla tutkittaviksi kollokaatioiksi vain sellaisia, joita esiintyy useammalla kuin yhdellä tekstintuottajalla.

Valaisen valittujen testien eroa vielä Stubbsin (1995a: 40) esimerkeillä: Stubbsin käyttämässä Cobuildin 120 miljoonan saneen korpuksessa verbi CAUSE ('aiheuttaa') esiintyi 308 kertaa ja substantiivi *trouble* ('harmi') 245 kertaa, yhteisesiintymä $f(\text{CAUSE}, \text{trouble}) = 8$. Kollokaatio saa *MI*-testin *I*-arvoksi 7,38, mikä on suhteellisen korkea ja korostaa semanttista yhteenkuuluvuutta, mutta *t*-testin arvoksi vain 2,81, mikä on taas matala ja johtuu siitä, että yhteisesiintymien määrä on pieni. Vastaavasti sanojen *the* ja *happy* ('onnellinen') vertailu antaa tuloksen: $f(\text{the}) = 80\ 833$, $f(\text{happy}) = 201$, $f(\text{the}, \text{happy}) = 33$, josta $I = 1,68$ ja $t = 3,95$. Tulokset korostavat nyt semanttista riippumattomuutta (*I*-arvo) ja kieliopillista suhteen voimakkuutta (*t*-arvo).

Tilastollisia menetelmiä vertaillen Stubbs (1995a: 40) on päätenyt esittämään molempien edellä kuvattujen testausapojen yhdistämiseksi seuraavanlaisen kolmiportaisen menetelmän, joka antaa tulokseksi sellaiset kollokaatiot, joiden esiintyminen ei ole todennäköisesti kiinni sattumasta:

- a) yhteisesiintymien yleisyysjärjestyksen listaaminen; vähimmäisesiintymärajan alittavien hylkääminen
- b) järjestäminen *MI*-testin mukaiseen järjestykseen; raja-arvon alittavien kollokaatioiden hylkääminen
- c) järjestäminen *t*-testin mukaiseen järjestykseen; raja-arvon alittavien kollokaatioiden hylkääminen.

Sovellan tätä menetelmää omassa työssäni lisäämällä siihen yhden portaan, jolla estetään idiosynkrasiasta johtuvien kollokaattien pääsy analyysiin (kohta a). Toisaalta Stubbsista poiketen sellaiset kollokaatiot, joiden *MI*-testin raja-arvo ei ylity mutta *t*-testin ylittyy, otetaan mukaan analyysiin. Näin *MI*-testin osuus ei painotu liiaksi "suodatuksessa". Alla olevassa portaikossa on mainittu myös käytettävät raja-arvot:

- a) yhteisesiintymien listaaminen; hylätään tapaukset, jotka esiintyvät vain yhdellä tekstintuottajalla
- b) yhteisesiintymien yleisyysjärjestyksen listaaminen; hylätään tapaukset, joissa $f(x,y) < 5$
- c) järjestäminen *MI*-testin mukaiseen järjestykseen; hylätään tapaukset, joissa *I*-arvo $< 3,00$, PAITSI jos *t*-testin raja-arvo ylittyy⁴⁴
- d) järjestäminen *t*-testin mukaiseen järjestykseen; hylätään tapaukset, joissa *t*-arvo $< 2,00$.

Muiden kuin kollokationaalisten valintojen testauksessa on käytetty pääasiassa suhteellisten osuuksien testaukseen tarkoitettua kahden riippumattoman otoksen *z*-testiä (ks. Pitkänen ja Kohonen 1984: 163; Butler 1985: 92–95) ja tarpeen vaatiessa myös muita tilastollisia menetelmiä.

5.4.3. Hakutermit, kotekstin analyysi ja tarkastelualue

Hakumenetelmät ja koodaus. Tutkimuksessa käytetty aineisto on lähes merkitsemätöntä juoksevaa tekstiä. Ainoa aineistossa oleva merkintä on tekstejä yksilöivä koodi, joka noudattaa Käännösuomen korpuksen alkuperäistä koodausta. Tätä koodausta tarvitaan tutkimuksessa silloin, kun on olennaista tietää, mistä tekstistä esiintymä on peräisin. Muulla tavalla tekstiä ei ole merkitty tunnistein eikä aineistoa ole myöskään lemmattu. Aineiston analysoimisessa on käytetty *WordSmith Tools*-ohjelmapakettia (Scott 1998), erityisesti sen *Concord*- ja *WordList*-ohjelmia. Näistä ensimmäinen tuottaa haetun hakusanan ja sen lähikontekstin esittävän konkordanssilistan eli niin sanotun KWIC (Key Word In Context) -konkordanssin. Ohjelma antaa mahdollisuuden muokata hakutulosta muun muassa oikean- ja vasemmanpuoleisen kotekstin mukaan esimerkiksi aakkojärjestyksen tai itse annetun koodauksen perusteella. Näitä toimintoja on käytetty tässä työssä runsaasti, koska niiden avulla lähikoteksti on voitu luokitella sanaluokkajaon, merkityspiirteiden ja yksittäisten sanojenkin perusteella. *WordList*- eli sanalistaohjelmaa on käytetty kollokationaalisten suhteiden analyysissa, jolloin on jouduttu laskemaan kollokaatteina tavattavien lekseemien esiintyminen korpuksessa. Koska aineistoa ei ole valmiiksi lemmattu, on haettavat kollokaatit lemmattu manuaalisesti tätä ohjelmaa käyttäen.

Havaintoyksiköt on haettu konkordansserilla käyttäen hakutermeinä merkkijonoja *hyvi**, *kovi** ja *oike**. Hakutermeissä käytetty *-korvausmerkki korvaa kyseiseltä paikalta yhden, useamman tai ei yhtään merkkiä. Tällä hakumenetelmällä saadaan hakutulosjoukkoon sellaisetkin muodot, jotka eivät ole yleiskielen mukaisia (esim. *oikeen*) tai jotka sisältävät mahdollisesti liitteitä. Hakutulos tuottaa kuitenkin myös runsaasti niin sanottua roskaa eli osumia, jotka eivät kuulu haluttuun para-

⁴⁴ *MI*-testin raja-arvona on käytetty toisinaan myös alemmaa arvoa: Barnbrook (1996: 99) on asettanut rajaksi $I = 1,58$, ks. myös Jantunen 2001b.

digmaan, koska hakutulosjoukkoon tulee mukaan myös esimerkiksi *hyvä-* ja *kova-* adjektiivien monikkomuotoja (esim. *hyvillä, koville*) ja *oikea-* adjektiivin taivutusmuotoja (esim. *oikea, oikeissa*) ja myös erilaisia yhdyssanoja (esim. *hyvinvointi, oikeuskansleri*). Tällaisten havaintoyksiköiden joukkoon kuulumattomien muotojen ja sanojen karsiminen onnistuu kuitenkin helposti konkordansserilla, joka luettelee tarvittaessa haun tuottamat hakusanat aakkosjärjestykseen, jolloin roskat ovat helposti poistettavissa. *Hyvin-* adverbien materiaaliset tapaa ilmoittavat käyttötavat (*Hän laulaa hyvin*) ja *hyvä-* adjektiivin instruktiivimuodot (*Hän valmistui hyvin arvosanoin*) on karsittu pois lähikotekstin avulla. Samoin on tehty *kova-* ja *oikea-* adjektiivien instruktiivi- ja positiivimuodoille ja *oikein-* adverbien modaalisille esiintymille (*Kirja on oikein mahdolloman hyvä*).

Hakuyksiköiden kolligationaalista valintaa tutkiessani olen keskittynyt leksemitason analyysiin (kolligaatioanalyysin mahdollisuuksista ks. lukua 2.3.3). Astemääritteiden liitepartikkelillisten muotojen kolligaatioanalyysia ei voi pitää järkevänä, koska esiintymien määrä on tähän liian pieni. Analyysin ulkopuolelle jää myös lauseketason kuvaus; määritellä ei ole myöskään semanttista variaatiota, joten eri merkitysten kolligaatioanalyysi on mahdoton. Kolligaatioanalyysi perustuu sanaluokkakolligaattien tutkimiseen, mutta näiden rinnalle olen ottanut luvussa 2.3.3 esitettyjen havaintojen pohjalta myös lauseenasemakolligaatin, joka konkretisoituu lauseenalkuisuutena ja -loppuisuutena. Koska kaunokirjallisille teksteille tyypillisten polveilevien ja monimutkaisten konstituenttirakenteiden sekä esimerkiksi elliptisten ilmausten vuoksi lauserajan ratkaiseminen ei osoittautunut läheskään aina yksinkertaiseksi, olen noudattanut koodauksessa A. Hakulisen ym. 1996 (1980) kvantitatiivisessa korpustutkimuksessa käyttämää luokittelua. Peruslähtökohtana pidän sitä, että lauseessa on finiittimuotoinen verbi, mutta A. Hakuliseen ym. (mts.10–20) nojautuen olen laskenut lausekonstruktioksi lisäksi seuraavat rakenteet (esimerkit pääosin A. Hakulinen ym. mp.):

- a) objektina oleva partisiippirakenne
Huomasin lasta työnnettävän vaunuissa.
- b) temporaalirakenne
Hiiret juoksevat pöydällä kissan ollessa poissa.
- c) elliptinen rinnasteinen finiittiverbirakenne
Kalle söi / ja lähti.
- d) elliptinen rinnasteinen finiittiverbitön rakenne, jos siinä on vähintään kaksi konstituenttia
Se syö leivänmuruja / ja Ø Ø nimensä mukaisesti sokeria.
- e) elliptinen vaihtoehdon sisältävä tai polaarisuudeltaan vastakohtainen rakenne
-- *soittaako Piirpauke omilla ehdoillaan, / sävelmateriaalin ehdoilla / vai yleisön ehdoilla.*
-- *suurin osa siirtomaissa tehdyistä investoinneista oli lähtöisin emämaan valtion kassasta / eikä Ø Ø siirtomaassa muodostetusta yksityisestä pääomasta.*
- f) subjektina oleva 1 infiniitti määritteineen
On väärin / luulla esitystä -- suorasukaisesti poleemiseksi.

Lisäksi olen käsitellyt lauseen tapaisina sellaisia elliptisiä rakenteita, joissa ei ole finiittiverbiä, mutta jotka on välimerkityksellä osoitettu selvästi omaksi lausemaiseksi kokonaisuudekseen. Tällainen on esimerkiksi aineistosta peräisin oleva tapaus *Tyttökö se on? - Niin, ja hyvin erikoinen lapsi*.

Morfologisen preferenssin analyysiin tarvittava morfologinen koodaus on tehty Connexorin *Machineese Phrase Tagger* -ohjelmalla⁴⁵. Ohjelma tuottaa monipuolisen analyysin, josta ilmenevät muun muassa sanojen perusmuodot, morfologinen koodaus sekä sanojen syntaktiset suhteet. Perusmuotojen ja syntaktisten suhteiden tarkistamisella on pystytty poistamaan ohjelman tekemät virheelliset tulkinat. Aineistosta poimitusta virkkeestä *Minä pyydän sinulta hyvin suurta palvelusta, oli Helena sanonut* ohjelma antaa taulukon 5.4 kaltaisen tuloksen:

Taulukko 5.4. Esimerkki Connexorin *Machineese Phrase Tagger* -ohjelmalla tehdystä morfologisesta analyysistä.

syötös	lekseemin perusmuoto	syntaktinen koodaus	morfologinen koodaus
Minä	minä	subj:>2	NH PRON SG1 NOM
pyydän	pyytää	main:>0	+MV V ACT IND PRES SG1
sinulta	sinä	sou:>2	NH PRON SG2 ABL
hyvin	hyvin	ad:>5	AD> ADV
suurta	suuri	attr:>6	A> A SG PTV
palvelusta	palvelus	obj:>2	NH N SG PTV
oli	olla		+MV V ACT IND PAST SG3
Helena	helena	subj:>8	NH N SG NOM
sanonut	sanoa	comp:>8	-MV V ACT PCP2 SG
.	.		
<s>	<s>		

Tarkastelualue. Hakusanaa eli noodia ympäröivän tarkastelualueen pituuden määrittäminen on yksi tärkeä tuloksiin vaikuttava tekijä itse aineiston valinnan ohella, koska tarkasteltavaan alueeseen kuuluvien sanojen on määrä kuvata hakusanan kontekstuaalisia ominaispiirteitä, ja sen ulkopuolella olevilla sanoilla taas ei katsota olevan samankaltaista sidoksisuutta hakusanaan (ks. mm. Berry-Rogghe 1973). Näin ollen tutkimukseen mukaan tulevien kollokaattien ja muiden kontekstuaalisten ominaisuuksien määrä rajautuu nimenomaan tarkastelualueen määrittämisen mukaan. Optimaaliseksi katsottu tarkastelualue on voinut olla hyvinkin eripituinen eri tapauksissa, kuten luvussa 2.3.2 tuli ilmi. Millainen olisi sitten käsillä olevaan tutkimukseen sopiva tarkastelualue? Riittääkö siihen vieruskollokaattien tai pelkästään jopa oikeanpuoleisten vieruskollokaattien eli astemääritteiden pääsanojen analysoiminen? Muun muassa Paradis (1997) on rajoittanut tutkimuksensa O1-kollokaatteihin, ja näissä pelkästään adjektiiviveihin, Bäcklund (1973) ja Klein (1998) tarkastelevat laajemmin kaikkia pääsanaluokkia (eli kolligaatteja) tässä positiossa.

⁴⁵ Ohjelman esittelyversio oli tutkimushetkellä vapaasti käytettävissä Internetissä osoitteessa www.connexor.com.

Kyseisissä tutkimuksissa on näinkin suppealla tarkastelualueella kyetty luomaan kuva käsiteltyjen astemääritteiden kontekstuaalisista ominaisuuksista. Pitäytyminen pelkästään pääsanapositiona rajoittaa tarkastelun kuitenkin vain astemääritteen ja sen pääsanojen kuvaukseen ja jättää ulkopuolelle muun kontekstin. Lisäksi pitäytyminen suppeassa tarkastelualueessa on omiaan korostamaan pelkästään vakiintuneita konstruktioita, kuten idiomeja, kun taas laaja tarkastelualue antaa mahdollisuuden laajemmalta alueelta esille tulevien semanttisten ja muiden preferenssien tutkimiseen (Church ja Hanks 1990: 23).

Vaikka astemääritteen ja sen pääsanojen muodostamat rakenteet ovatkin erittäin relevantteja kuvattaessa yhdistymismahdollisuuksia (ks. esim. Orpana 1988), saattaa myös tämän rakenteen ylittävä konteksti tuoda esiin muita astemääritteille tyypillisiä ominaisuuksia, ja etenkin käsillä olevan työn kannalta olennaisia piirteitä niin synonymian kuin kielimuotojenkin suhteen. Jos kuvauksen kohteena ovat pelkästään O1-position lekseemit, ulkopuolelle jäävät esimerkiksi O2-position ilmaukset, jotka voivat olla pääsanoina niissä lausekkeissa, joissa astemääritteitä on käytetty, kuten *päivä* nominilausekkeessa *hyvin kaunis päivä*. Myös vasemmanpuoleisen kontekstin huomiotta jättäminen kaventaisi kuvausta: silloin kuvauksen ulkopuolelle jäisivät predikaattiverbit sekä subjektit, jotka edeltävät esimerkiksi predikaatiivirakenteita kuten virkkeessä *Päivä oli hyvin kaunis*. Jotta kuvauksessa tavoitetaan tarkempi kuva astemääritteiden kontekstuaalisista ominaisuuksista, tarkastelualueetta on tarpeen laajentaa oikeanpuoleista vieruskollokaattia laajemmaksi. Koska tarkastelualue on varioinut havaintoyksikön ja tutkimustavoitteiden mukaan runsaastikin, ei aikaisemmista tutkimuksista löydy sellaisenaan valmista havaintoaluetta, jota voitaisiin käyttää nykyisessä tutkimuksessa. Olen tavoitellut tutkimuksen päämäärille sopivaa tarkastelualueita aineistopohjaisella menetelmällä, jonka kuvaan seuraavassa.

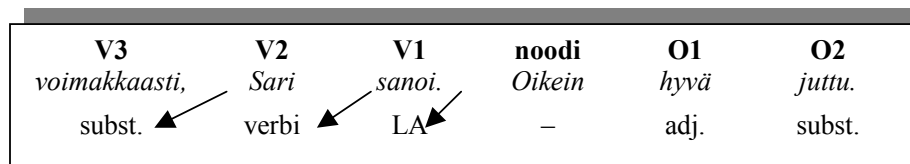
Tutkimuksessa käytettävän tarkastelualueen rajaamiseksi olen tehnyt testin, joka perustuu astemääritteiden konteksteissä olevien kieliopillisten rakenteiden analyysiin. Aluetta ei valita siis ulkopuolisella määrittelyllä, vaan testit antavat aineistoon perustuvan ehdotuksen tarkastelualueen laajuudesta. Vaikka Hallidayn (1966: 151–152) mielestä tarkastelualueen rajoja ei voisi rajata tyydyttävästi kieliopillisin perustein, olen kuitenkin päätenyt tähän ratkaisuun, koska kontekstuaalisten valintapiirteiden analyysissä on mukana kieliopillisten kategorioiden kuten sanaluokkien analyysi kolligaatioanalyysissä ja morfologisten luokkien analyysi, joiden tuloksiin vaikuttavat ratkaisevasti ne kieliopilliset rakenteet, jotka havaintoyksiköitä ympäröivät, ja siten myös havaintoalueen koko. Havaintoyksiköiden kontekstin rakenteita etukäteen tarkastelemalla voidaan tarkastelualue määrittää paremmin kuin vain etukäteen valitsemalla jokin, esim. intuitioon tai aiempiin tutkimuksiin perustuva alue.

Rakenteita olen tarkastellut määritteen molemmilta puolilta. Oikeanpuoleisesta kontekstista olen ottanut huomioon rakenteet, joissa astemäärite on määritteenä. Näin esimerkiksi lauseessa *Tänään on hyvin kaunis päivä* on kaksi oikeanpuoleista sanaa, joihin astemäärite rakenteellisesti liittyy ja lauseessa *Tänään on hyvin sateinen ja tuulinen päivä* neljä sanaa. Toisaalta vasemmanpuoleisesta kontekstista on laskettu subjekti-, objekti-, predikaatti- ja adverbiaalirakenteet, jotka ilmaisevat muun muassa ”kuka tai mikä on *hyvin* jotakin” tai ”milloin joku tai jokin tekee jotakin

hyvin jollakin tavalla”. Niinpä lause *Eilen Alma lauloi kirkossa hyvin kauniisti* antaa vasemmanpuoleisten mukaan laskettavien saneiden määräksi neljä.

Varsinaiseen tarkastelualueeseen on päädytty tekemällä yksi konkordanssi kaikista kolmen astemäärittteen esiintymistä alkuperäissuomen korpuksessa, valitsemalla tästä systemaattisen otannan avulla ensimmäinen ja siitä edelleen joka 20:s rivi, jolloin päädyttiin 33 konkordanssiriviin. Näistä riveistä on laskettu edellisen kuvauksen perusteella sanemäärät ja edelleen näiden keskiarvot hakusanan molemmiin puolin. Testin mukaan vasemmanpuoleinen koteksti koostuu keskimäärin 3,4 saneen pituisista rakenteista ja oikeanpuoleinen 2,0 saneen rakenteista. Tulokset tarkistettiin toistamalla testi siten, että saman konkordanssilistan riveistä otettiin toinen konkordanssirivi ja siitä edelleen joka 20:s. Tämä testi antoi tulokseksi lukemat 3,09 ja 1,94. Näiden keskiarvolukujen perusteella tämän työn tarkastelualueeksi on valittu 3 lekseemiä hakusanan vasemmalta puolelta ja 2 sanaa oikeanpuoleisesta kotekstista eli alue V3–O2, joka on kuvattu kuviossa 5.2. Alue kattaa siten vasemmalta puolelta yhden kollokaatin oikeanpuoleista enemmän eli on epäsymmetrinen.

Kollokaatit ja muut kontekstuaaliset piirteet lasketaan piittaamatta esimerkiksi puhujan, kertojan tai kerrontatavan vaihdoksesta, ja kollokaattien kytkennöissä ei ole otettu huomioon tarkastelualueelle sattuvaa lauserajaa. Tarkastelualueita on kuitenkin paikoin muutettava toisaalta tavoitteiden ja toisaalta analyysin mielekkyyden kannalta. Esimerkiksi kolligaattianalyysin yhteydessä kotekstilauseita analysoidessa aluetta venytetään lauseen mittaiseksi (siltoin kun lause ei ole samankoinen tai pienempi kuin tarkastelualue), koska lausepituus osoittautuu yhdeksi tärkeäksi muuttujaksi. Tarkastelualueita joudutaan kasvattamaan isommaksi pakosta myös semanttista prosodiaa analysoidessa, sillä prosodisten ominaisuuksien hahmottamiseen tarvitaan toisinaan hyvinkin laaja koteksti. Toisaalta semanttisen preferenssin tarkastelussa on järkevää keskittyä vain O1-positioon koko tarkastelualueen kollokaattien runsaan semanttisen kirjon vuoksi (tarkemmin luvussa 6.4.1).



Kuvio 5.2. Aineiston käsittelyyn valittu tarkastelualue, alemmalla rivillä esimerkkinä kolligaattianalyysi (V = vasen, O = oikea).

6. SYNONYIMIT *HYVIN*, *KOVIN* JA *OIKEIN* KONTEKSTUAALISINA MERKITYSYKSIKKÖINÄ

Synonyymisyyttä ja sitä rikkovia kontekstuaalisia merkityspiirteitä käsitellään seuraavassa neljässä eri alaluvussa astemääritteiden *hyvin*, *kovin* ja *oikein* avulla: ensimmäinen keskittyy niiden kollokationaalisiin ja toinen kolligationaalisiin valintoihin, kolmas morfologiseen preferenssiin ja neljäs semanttisiin valintoihin. Viimeisessä alaluvussa luodaan kustakin astemääritteestä aikaisemmissa luvuissa esitettyjen tulosten perusteella kontekstuaalinen profiili, joka kuvaa astemääritteitä kontekstuaalisina merkitysyksikköinä. Tässä luvussa esitetty astemääritteiden kontekstuaalisten merkitysten ja synonyymisyyden kuvaus perustuu alkuperäissuomen eli AS-aineistoon.

6.1. Kollokationaalinen valinta

Astemääritteiden kollokationaalisten valintojen kuvaus etenee tilastollisesti merkitsevien ja frekventtien kollokaattien käsittelystä kollokationaalisen diversiteetin ja viimein astemääritteiden ja niiden kollokaattien muodostamien klustereiden tarkasteluun. Lopuksi luodaan yhteenvedoksi profiili kunkin astemääritteen kollokationaalisisista ominaisuuksista.

6.1.1. Kollokationaalinen valinta tarkastelualueella

6.1.1.1. Kollokaattien tilastollinen merkitsevyys

Synonyymisten astemääritteiden kollokationaalista valintaa tarkastellaan tilastollisin menetelmin käyttäen laskennassa luvussa 5 kuvatuin perustein kahta merkitsevyydestiä, *MI*-testiä ja *t*-testiä. Astemääritteitä on aineistossa seuraavasti: *hyvin* 352, *kovin* 176 ja *oikein* 121. Tarkastelualueen V3–02 koko on saneissa ilmoitettuna *hyvin*-määritteellä 1760 (5x352), *kovin*-määritteellä 880 (5x176) ja *oikein*-määritteellä 605 (5x121) sanetta. Tarkastelualueella olevien vähintään viisi kertaa esiintyvien kollokaattien tilastolliset merkitsevyydet on esitetty taulukoissa 6.1, 6.2 ja 6.3, joihin on listattu merkitsevyydsarvojen lisäksi myös kollokaatin sanemäärä korpuksessa ja odotettu frekvenssi tarkastelualueella. Raja-arvot (*MI*-testissä 3,00 ja *t*-testissä 2,00) ylittävät merkitsevyydsarvot ja kollokaatit on taulukoissa lihavoitu. Taulukoista on luettavissa kollokaatit, jotka ovat merkitseviä molempien testien perusteella, ja kollokaatit, jotka osoittautuvat merkitseviksi vain *t*-testin perusteella.

Taulukko 6.1. *hyvin*-määritteen kollokaattien (min. 5 esiintymää) tilastollinen merkitsevyys.

kollokaatti	frekvenssi	f korpuksessa	odotettu f	<i>MI</i> -testi	<i>t</i> -testi
olla	158	52635	93,89	0,75	5,10
ja	83	39473	70,41	0,24	1,38
minä	46	17414	31,06	0,57	2,20
se	43	26212	46,76	-0,12	-0,57
hän	41	17152	30,60	0,42	1,62
hyvin	17	352	0,63	4,76	3,97
sanoa	17	7599	13,56	0,33	0,84
mutta	16	6159	10,99	0,54	1,25
että	14	11858	21,15	-0,60	-1,91
jokin	13	2338	4,17	1,64	2,45
sinä	13	4723	8,43	0,63	1,27
hitaasti	12	280	0,50	4,59	3,32
niin	12	5962	10,64	0,17	0,39
tämä	10	4903	8,75	0,19	0,40
ei	9	19740	35,21	-1,97	-8,74
mies	9	3546	6,33	0,51	0,89
tehdä	9	2870	5,12	0,81	1,29
tulla	9	4770	8,51	0,08	0,16
vain	9	2421	4,32	1,06	1,56
ihminen	8	2170	3,87	1,05	1,46
joka	8	9495	16,94	-1,08	-3,16
pieni	8	1125	2,01	2,00	2,12
vähän	8	648	1,16	2,79	2,42
hiljaa	7	279	0,50	3,81	2,46
sitten	7	2745	4,90	0,52	0,80
vanha	7	1315	2,35	1,58	1,76
väsinyt	7	85	0,15	5,53	2,59
katsoa	6	2120	3,78	0,67	0,91
kysyä	6	1965	3,51	0,78	1,02
nainen	6	2287	4,08	0,56	0,78
nyt	6	3122	5,57	0,11	0,18
näyttää	6	1086	1,94	1,63	1,66
pian	6	296	0,53	3,51	2,23
alkaa	5	1939	3,46	0,53	0,69
kaunis	5	442	0,79	2,66	1,88
moni	5	909	1,62	1,62	1,51
nuori	5	905	1,61	1,63	1,51
paljon	5	925	1,65	1,60	1,50
vaikka	5	1796	3,20	0,64	0,80
varovasti	5	129	0,23	4,44	2,13
voida	5	2926	5,22	-0,06	-0,10

Koska *MI*-testi nostaa merkitseviksi yhteisesiintymät, joiden jäsenten frekvenssit saattavat olla aineistossa pieniä mutta jotka silti esiintyvät yhdessä suhteellisen yleisesti tarkastelualueella, merkitseviksi erottuvat nytkin juuri sanat, joiden frekvenssi on korpuksessa melko vähäinen: suurimmillaan frekvenssi on kollokaatilla *hyvin* (352 esiintymää) ja pienimmillään kollokaatilla *VÄSYNYT* (85 esiintymää). Pelkästään *t*-testin merkitseviksi nostamia ovat taajimmin esiintyvä *OLLA*, frekventti pronomini *MINÄ* ja kolme muuta suhteellisen taajasti esiintyvää sanaa. Merkitsevien kollokaattien joukosta voinee merkitsevimpinä pitää sellaisia kollokaatteja, jotka nousevat esiin molemmissa testeissä. Näin ollen *hyvin*-määritteen kollokaateista osoittautuvat merkitsevimmiksi *hyvin*, *hitaasti*, *hiljaa*, *VÄSYNYT*, *pian* ja *varovasti*. Lisäksi *t*-testi nostaa merkitseviksi myös kollokaatit *OLLA*, *MINÄ*, *JOKIN*, *PIENI* ja *vähän*. Huomion arvoista on, että molempien testien perusteella merkitseviksi osoittautuvat ovat sanoja, jotka ovat lähes yksinomaan positiossa O1, eli sanoja, joita *hyvin* määrittää. Tällaisia ovat myös *PIENI* ja *vähän*, mutta muut *t*-testin esiin nostamista kollokaateista ovat asemaltaan varioivia. Taulukosta voidaan poimia myös testien vähiten merkitseviksi osoittamat kollokaatit, joita ovat *EI*, *JOKA*, *SE* ja *ETTÄ*. Näiden osalta testit antavat samanlaisen tuloksen.

kovin-määritteelle testaus antaa taulukon 6.2 mukaisen tuloksen. Tilastollisesti merkitseviksi kollokaateiksi osoittautuvat molempien testien perusteella *paljon*, *MONI*, *usein* ja *TÄRKEÄ*, pelkästään *t*-testin mukaan merkitseviä ovat *OLLA*, *EI* ja *mutta*. Tässäkin on havaittavissa sama suuntaus kun edellä: molempien testien perusteella merkitseviä ovat määritettävänä olevat O1-kollokaatit, kun taas *t*-testin merkitseviksi osoittamat ovat jälleen kollokaatteja, jotka esiintyvät useammassa positiossa. Vähiten merkitseviä ovat *ETTÄ*, *ja*, *HÄN* sekä *MINÄ*.

Taulukkoon 6.3 on lueteltu vuorostaan *oikein*-määritteen frekventimpien kollokaattien tilastolliset merkitsevyydet. Molempien testien perusteella merkitseviksi kollokaateiksi erottautuvat *hyvin*, *HYVÄ*, *ISO*, *TUNTUA*; *t*-testin perusteella merkitsevä on myös *jos*. Vähiten merkitseviä ovat vuorostaan *HÄN*, *ja*, *EI* sekä *OLLA*. Tulos vahvistaa aiemmin huomattua: molempien testien perusteella merkitseviksi nousevat ovat määrittäyssuhteen pääsanoja – paitsi kollokaatti *TUNTUA*. Vertailun helpottamiseksi on merkitseviksi osoittautuneet kollokaatit lueteltu kertaalleen taulukkoon 6.4.

Synonyymisten astemääritteiden kollokaatiiviset valinnat osoittautuvat selvästi erilaisiksi: *hyvin* ja *kovin* jakavat vain yhden kollokaatin (*OLLA*) ja samoin myös *hyvin* ja *oikein* (*hyvin*). Jälkimmäisiä yhdistävä *hyvin*-kollokaatti on kuitenkin kontekstissa eri tehtävissä, kuten ilmenee myöhemmin konkordanssista 6.6 (s. 96) adverbikollokaattien käsittelyn yhteydessä. *kovin*- ja *oikein*-määritteillä ei vuorostaan ole ainuttakaan yhteistä tilastollisesti merkitsevää kollokaattia.

Merkitseviksi osoittautuneiden kollokaattien lisäksi eroja on määritteiden kesken myös siinä, mitkä kollokaatit osoittautuvat vähiten merkitseviksi eli negatiivisiksi kollokaateiksi. Erityisesti huomio kiinnittyy kollokaattiin *EI*, joka osoittautui *kovin*-määritteen merkitseväksi kollokaatiksi, mutta muilla astemääritteillä se on vähiten merkitsevien kollokaattien joukossa, *hyvin*-määritteillä se on jopa vähiten merkitsevä molempien testien mukaan (taulukko 6.1). Täten *EI*-kollokaatti osoittautuu erittäin selvästi yhdeksi astemääritteiden synonyymisyyttä rikkovaksi tekijäksi. *OLLA* on merkitsevä sekä *hyvin*- että *kovin*-määritteillä, mutta ei *oikein*-määritteillä.

Taulukko 6.2. *kovin*-määritteen kollokaattien (min. 5 esiintymää) tilastollinen merkitsevyys.

kollokaatti	frekvenssi	f korpuksessa	odotettu f	MI-testi	t-testi
olla	90	52635	46,95	0,94	4,54
ei	82	19740	17,61	2,22	7,11
se	31	26212	23,38	0,41	1,37
hän	18	17152	15,30	0,23	0,64
mutta	16	6159	5,49	1,54	2,63
ja	15	39473	35,21	-1,23	-5,22
minä	13	17414	15,53	-0,26	-0,70
paljon	13	925	0,83	3,98	3,38
että	11	11858	10,58	0,06	0,13
moni	11	909	0,81	3,76	3,07
niin	8	5962	5,32	0,59	0,95
usein	7	253	0,23	4,96	2,56
he	5	3007	2,68	0,90	1,04
mies	5	3546	3,16	0,67	0,82
nainen	5	2287	2,04	1,29	1,32
tämä	5	4903	4,37	0,19	0,28
tärkeä	5	243	0,22	4,53	2,14
voida	5	2926	2,61	0,94	1,07
vuosi	5	1237	1,10	2,18	1,74

Taulukko 6.3. *oikein*-määritteen kollokaattien (min. 5 esiintymää) tilastollinen merkitsevyys.

kollokaatti	frekvenssi	f korpuksessa	odotettu f	MI-testi	t-testi
olla	29	52635	32,28	-0,15	-0,61
hyvin	27	562	0,34	6,29	5,13
hyvä	21	1556	0,95	4,46	4,37
ja	18	39473	24,20	-0,43	-1,46
se	17	26212	16,07	-0,01	0,22
minä	15	17414	10,68	0,49	1,12
jos	9	2588	1,59	2,50	2,47
ei	8	19740	12,10	-0,60	-1,45
että	7	11858	7,27	-0,05	-0,10
mutta	6	6159	3,78	0,67	0,91
sanoa	6	7599	4,66	0,36	0,55
tämä	6	4903	3,01	1,00	1,22
hän	5	17152	10,52	-1,07	-2,47
iso	5	693	0,42	3,56	2,05
kaikki	5	3173	1,95	1,36	1,37
nyt	5	3122	1,91	1,39	1,38
sinä	5	4723	2,90	0,79	0,94
sitten	5	2745	1,68	1,57	1,48
tuntua	5	767	0,47	3,41	2,03

Taulukko 6.4. Tilastollisesti merkitsevät kollokaatit astemääritteittäin.

<i>hyvin</i>	<i>kovin</i>	<i>oikein</i>
hiljaa	ei	hyvin
hitaasti	moni	hyvä
hyvin	mutta	iso
jokin	olla	jos
minä	paljon	tuntua
olla	tärkeä	
pian	usein	
pieni		
varovasti		
vähän		
väsynyt		

Myöhemmin taulukosta 6.5 (s. 91) käykin ilmi, että *oikein*-määritteellä on OLLA-verbistä verbikollokaatteina selvästi synonyymejaan vähemmän, niiden osuus on vain 4,8 (n = 29) prosenttia kun muilla osuus on 9,0 (n = 158) ja 10,2 (n = 90). Tilastollinen testaus osoittaa OLLA-verbistä *oikein*-määritteen negatiiviseksi kollokaatiksi. MINÄ taas erottaa määritteet *hyvin* ja *kovin* selvästi toisistaan: siinä missä MINÄ on yksi merkitsevimmistä *hyvin*-määritteen kollokaateista, on se *kovin*-määritteen kollokaateista toiseksi vähiten merkitsevä.

Mielenkiintoisella tavalla taulukossa 6.4 luetelluista kollokaateista erottuu semanttisia vastakohtapareja, jotka myös erottavat näitä synonyymeja toisistaan. Kun *hyvin* kollokoi PIENI-adjektiivin kanssa, *oikein* saa kollokaatikseen ISO-adjektiivin. Vastaavasti *hyvin* <vähän> -kollokaatiota vastaa antonyyminen *kovin* <paljon>. Tällaiset erot antavat aiheen olettaa, että synonyymisilla määritteillä saattaisi olla omanlaisiaan semanttisia preferenssejä. (Semanttista preferenssiä käsitellään myöhemmin luvussa 6.4.) Tilastollisen testauksen perusteella synonyymisillä astemääritteillä näyttää siten olevan omanlaisensa kollokationaaliset valinnat. Valintoja tarkastellaan vielä yksityiskohtaisemmin frekvenssianalyysin avulla.

6.1.1.2. Kollokaattien esiintymistäajuus

Edellinen katsaus perustui kahden merkitsevyydestin avulla tehtyyn analyysiin ja osoitti synonyymit kontekstuaalisilta valinnoiltaan erilaisiksi. Tilastollisesti merkitseviksi osoittautuvat kollokaattien määrä oli kuitenkin – ehkä yllättävänkin – vähäinen. Esiin tullut tilanne johtuu mitä ilmeisimmin siitä, että kieliopillisia sanoja lähentyvien astemääritteiden kollokationaalinen valinta ei ole kovin voimakasta, kuten ei kieliopillisilla sanoilla yleensääkään. Tästäkin huolimatta selkeitäkin valintoja oli havaittavissa. Seuraavaksi luodaan kuitenkin katsaus kaikkiin vähintään viisi kertaa tarkastelualueella esiintyviin kollokaatteihin. Tarkoituksena on ensinnäkin täydentää tilastollisin menetelmin saatua kuvaa ja toiseksi esittää kollokationaalisen diversiteetin laskennassa käytettävät esiintymätaajuudet. Kaikki vähintään viidesti

esiintyvät kollokaatit on lueteltu taulukkoon 6.5 esiintymistaajuuksineen ja suhteellisine osuuksineen, viimeiset on laskettu koko tarkastelualueen kaikkien kollokaattien määrästä.

Kollokaatteja, joita tavataan vähintään kahdella tekstintuottajalla ja jotka esiintyvät vähintään viidesti, on astemääritteillä *kovin* ja *oikein* saman verran (19 kappaletta) mutta *hyvin*-määritteellä yli kaksinkertainen määrä, 41 kappaletta. Jälkimmäisen määritteen vähintään viisi kertaa esiintyvien kollokaattien suurempi määrä johtuu suurelta osin siitä, että *hyvin* on esiintymistaajuudeltaankin selvästi synonyymejaan yleisempi, mikä mahdollistaa sen, että vähintään viisi kertaa esiintyviä kollokaattejakin on enemmän. Siten kollokaattien *lukumäärän* vertailu ei ole mielekäästä. Astemääritteiden kollokationaalista yhtäläisyyttä ja variaatiota voidaan sen sijaan analysoida tarkastelemalla, millaisia kollokaatit ovat, kuinka yleisiä yksittäiset kollokaatit ovat ja mikä on yleisimpien kollokaattien osuus.

Verbikollokaatit. Jokaisen astemääritteen yleisin kollokaatti on OLLA. Tätä ei voi pitää yllättävänä, onhan OLLA suomen kielen yleisin sana (Saukkonen ym. 1979: 41; Heikkinen ym. 2001: 15) ja siten löydettävissä yleensä aina kollokaattilistojen kärkipäästä. Taulukossa on joka kollokaattisarakeessa myös EI, mutta sen frekvenssi ja suhteellinen osuus vaihtelee huomattavasti – tämä verbikollokaattihan erotti astemääritteet toisistaan myös tilastollisessa testauksessa. Näiden lisäksi astemääritteiden kotekstissa on toki muitakin verbikollokaatteja, ja muutamat ylittävät neljän esiintymän rajan. Synonyymien yleisimpien verbien kirjo näyttää aineiston valossa seuraavalta (prosenttiluku kertoo yksittäisen kollokaatin osuuden kaikista samaan ryhmään kuuluvista kollokaateista):

hyvin <OLLA 41 %, SANOA 4 %, EI, KATSOA, KYSYÄ, NÄYTTÄÄ, TEHDÄ, TULLA 2 %, ALKAA, VOIDA 1 %>

kovin <OLLA 31 %, EI 28 %, VOIDA 2 %>

oikein <OLLA 20 %, EI 5 %, SANOA 4 %, TUNTUA 3 %>

Maininnan arvoista on, että vaikka *hyvin*-määritteen verbikollokaattien kirjo näyttää muita suuremmalta, on muistettava, että itse määritekin esiintyy aineistossa muita useammin. Näin ollen määritteiden verbikollokaattien todellisesta variaatiosta ei voi tehdä pitäviä päätelmiä edellä olevan luettelon perusteella. Se antaa kuitenkin mahdollisuuden muunlaisiin päätelmiin. Ensinnäkin on havaittavissa, että OLLA ja EI yhdistävät kaikkia synonyymeja, mutta vain OLLA on kaikilla taajaan esiintyvä. Sen osuus näyttää kuitenkin vaihtelevan selvästi astemääritteittäin, ja erot ovat myös tilastollisesti merkitseviä⁴⁶. Toiseksi EI on huomattavasti yleisempi verbikollokaatti *kovin*-määritteellä kuin muilla: EI on sen kollokaattina lähes yhtä usein kuin OLLA, ja erot synonyymien osuuksiin ovat tilastollisesti merkitseviä⁴⁷. *kovin*-määritteellä on siis voimakas preferenssi kieltoverbiin.

⁴⁶ *hyvin* vs. *kovin* $z = 2,70086$ ($p < 0,01$); *hyvin* vs. *oikein* $z = 4,646$ ($p < 0,001$); *kovin* vs. *oikein* $z = 2,525313$ ($p < 0,05$).

⁴⁷ *hyvin* vs. *kovin* $z = 9,74679$ ($p < 0,001$); *kovin* vs. *oikein* $z = 5,587292$ ($p < 0,001$).

Taulukko 6.5. Astemääritteiden *hyvin*, *kovin* ja *oikein* vähintään 5 kertaa esiintyvät kollokaatit yleisyysjärjestyksessä tarkastelualueella V3–O2. (f = frekvenssi, f % = suhteellinen frekvenssi kaikista kollokaateista.)

<i>hyvin</i>			<i>kovin</i>			<i>oikein</i>		
kollokaatti (yht. 1760)	f	f %	kollokaatti (yht. 880)	f	f %	kollokaatti (yht. 605)	f	f %
olla	158	9,0	olla	90	10,2	olla	29	4,8
ja	83	4,7	ei	82	9,3	hyvin	27	4,5
minä	46	2,6	se	31	3,5	hyvä	21	3,5
se	43	2,4	hän	18	2,0	ja	18	3,0
hän	41	2,3	mutta	16	1,8	se	17	2,8
hyvin	17	1,0	ja	15	1,7	minä	15	2,5
sanoa	17	1,0	minä	13	1,5	jos	9	1,5
mutta	16	0,9	paljon	13	1,5	ei	8	1,3
että	14	0,8	että	11	1,3	että	7	1,2
jokin	13	0,7	moni	11	1,3	mutta	6	1,0
sinä	13	0,7	niin	8	0,9	sanoa	6	1,0
hitaasti	12	0,7	usein	7	0,8	tämä	6	1,0
niin	12	0,7	he	5	0,6	hän	5	0,8
tämä	10	0,6	mies	5	0,6	iso	5	0,8
ei	9	0,5	nainen	5	0,6	kaikki	5	0,8
mies	9	0,5	tämä	5	0,6	nyt	5	0,8
tehdä	9	0,5	tärkeä	5	0,6	sinä	5	0,8
tulla	9	0,5	voida	5	0,6	sitten	5	0,8
vain	9	0,5	vuosi	5	0,6	tuntua	5	0,8
ihminen	8	0,5						
joka	8	0,5						
pieni	8	0,5						
vähän	8	0,5						
hiljaa	7	0,4						
sitten	7	0,4						
vanha	7	0,4						
väsinyt	7	0,4						
katsoa	6	0,3						
kysyä	6	0,3						
nainen	6	0,3						
nyt	6	0,3						
näyttää	6	0,3						
pian	6	0,3						
alkaa	5	0,3						
kaunis	5	0,3						
moni	5	0,3						
nuori	5	0,3						
paljon	5	0,3						
vaikka	5	0,3						
varovasti	5	0,3						
voida	5	0,3						

Preferenssi tulee esiin vielä selvemmin, kun tarkastellaan sitä, onko kieltoverbi samassa lauseessa kuin astemääräite (ja pääsana, johon kieltö yleensä kohdistuu). Koska osa kotekstissa olevista kieltoverbeistä on todennäköisesti ennalta asetetun tarkastelualueen ulkopuolella, päästään aluetta laajentamalla selville siitä, kuinka moni astemääräitteistä tosiasiaa esiintyy samassa lauseessa kieltoverbin kanssa. Tarkastelun tulos (taulukko 6.6) paljastaa *hyvin*- ja *oikein*-määritteet selvästi myöntöhakuisiksi ja *kovin*-määritteen vuorostaan kieltöhakuisiksi. Eri verbikollokaattien esiintymistä on lueteltu esimerkkejä konkordanssiin 6.1.

Taulukko 6.6. Astemääritteiden esiintyminen samassa lauseessa kieltoverbin kanssa.

	<i>hyvin</i>		<i>kovin</i>		<i>oikein</i>	
	f	f %	f	f %	f	f %
lauseessa on kieltoverbi	2	0,6	118	67,0	3	2,5
lauseessa ei ole kieltoverbiä	350	99,4	58	33,0	118	97,5
-----	-----	-----	-----	-----	-----	-----
yhteensä	352	100,0	176	100,0	121	100,0

Konkordanssi 6.1. Esimerkkejä verbikollokaateista ja niiden asemasta kotekstissa.

1	ntäneljävuotiaana minä olin vielä kerran	<i>hyvin</i>	nuori, hyvin nuori yhden vuoden v
2	iten me elämme. Minä voisin tehdä siitä	<i>hyvin</i>	kevyen, niin että sinun olisi helppo
3	olasta taksiin mutta kuten hän itse sanoi ,	<i>hyvin</i>	säännöllistä. Amputoinnin jälkeen
4	pään asentoon. Sitten hän kääntyi hyvin	<i>hyvin</i>	hitaasti katsomaan Louneria. Nais
5	kosta. Sitä ei onneksi ollu viime aikoina	<i>kovin</i>	moni nähny . Jos tuo kumma äijä il
6	ltakin Rhealle yhdentekevää. - Se ei ole	<i>kovin</i>	kaukana sinun nykyisestä toimesta
7	tapahtumaan, emme Jumalaan emmekä	<i>kovin</i>	paljon ihmiseenkään, ainakaan ihm
8	ei kai vanhan huoran luonnon päälle voi	<i>kovin</i>	paljon ottaa yks keikka, varsinki
9	stunut Sariin, ja Sari sanoi Topin olevan	<i>oikein</i>	hauska mies, suorastaan huvittava
10	elimeen: ”Voiksä tulla? Kullervo ei voi	<i>oikein</i>	hyvin.” Meillä jatkui Antin kanssa
11	än on raitis ja kunnollinen, ja se tuntuu	<i>oikein</i>	hyvältä sen jälkeen missä minäkin
12	ielä ja mitä se halusi . Portailla kuuntelin	<i>oikein</i>	pitkään. Oli aivan hiljasta. Kävelin

Konjunktiokollokaatit. Astemääritteiden kollokaattiluetteloiden kärjessä on edellä käsiteltyjen verbien lisäksi muita kielessä taajaan esiintyviä sanoja, tavallisimmin kieliopillisia sanoja kuten konjunktioita ja pronomineja⁴⁸. Konjunktiokollokaattien kirjo on melko samanlainen kaikilla synonyymeilla, mikä ei ole sinänsä yllättävää, sillä konjunktioiden luokka on rajattu ja variaation mahdollisuus on siten pieni. Sen

⁴⁸ Myös aiemmin käsitelty kieltoverbi *ei* voidaan laskea kieliopilliseksi sanaksi (Karvonen 1995: 189).

lisäksi että konjunktiokollokaattien sijoitus yleisimpien kollokaattien joukossa vaihtelee etenkin *ja*- sekä *mutta*-konjunktioiden osalta (taulukko 6.5), ovat yksittäisten konjunktioiden osuudet erilaisia myös määritteittäin:

hyvin <*ja* 60 %, *mutta* 12 %, *että* 10 %, *vaikka* 4 %>

kovin <*mutta* 33 %, *ja* 31 %, *että* 23 %>

oikein <*ja* 40 %, *jos* 20 %, *että* 16 %, *mutta* 13 %>

hyvin näyttää poikkeavan synonyymeistaan siinä, että yksi konjunktio eli rinnastuskonjunktio *ja* on sen kontekstissa huomattavasti frekventimpi (60 %, $n = 83$) kuin sen muut konjunktiokollokaatit. Näin suurta eroa seuraavaksi yleisimpään kollokaattiin ei kahdella muulla astemääritteellä ole havaittavissa. *hyvin*-määritteen runsas *ja*-konjunktion osuus on tilastollisesti merkitsevästi suurempi kuin *kovin*- ja *oikein*-määritteellä ja *mutta*-konjunktion osuus vuorostaan suurempi *kovin*-määritteellä sen synonyymeihin verrattuna⁴⁹. *mutta* osoittautui myös tilastollisen testauksen perusteella juuri *kovin*-määrittele tyypilliseksi (taulukko 6.2). *kovin* esiintyykin synonyymejaan useammin adversatiivissa lauseissa, joissa on rinnastuskonjunktio *mutta* ja jotka ilmaisevat jotain vastakohtaista edellä kerrottuun (ks. konkordanssi 6.2 rivi 6). Lisäksi *että*-konjunktion osuus on *kovin*-määritteellä *hyvin*-synonyymiaan suurempi⁵⁰, ja konjunktio *jos* on tyypillisesti vain *oikein*-määritteen kontekstissa taajaan esiintyvä konjunktiokollokaatti. Konjunktiokollokaatteja on löydettävissä tavallisimmin oikeanpuoleisesta kontekstista positiosta O2, mutta melko runsaanlaisesti myös vasemmanpuoleisesta kontekstista, esimerkkejä konkordanssissa 6.2.

Pronominikollokaatit. Konjunktioiden tapaan myös pronominiinien luokka on suljettu, eikä pronominiokollokaattien⁵¹ variaatio ole juurikaan erilainen astemääritteiden kesken: pronominiokollokaattien osuuksissa ei ole järin suurta eroa, eivätkä synonyymit osoita kovin selvää preferenssiä johonkin yksittäiseen pronominiin, joskin *kovin*-määritteen SE on selvästi sen muita pronominiokollokaatteja yleisempi (34 %, $n = 31$), kun muilla yleisimpien pronominiokollokaattien jakauma on tasaisempi. Esimerkkejä pronominiokollokaateista konkordanssissa 6.3.

hyvin <MINÄ 25 %, SE 24 %, HÄN 23 %, SINÄ 7 %, TÄMÄ 5 %, JOKA 4 %>

kovin <SE 34 %, HÄN 20 %, MINÄ 14 %, HE, TÄMÄ 6 %>

oikein <SE 30 %, MINÄ 27 %, TÄMÄ 11 %, HÄN, SINÄ 9 %>

⁴⁹ *ja*: *hyvin* vs. *kovin* $z = 3,404158$ ($p < 0,001$); *hyvin* vs. *oikein* $z = 2,309716$ ($p < 0,05$).

mutta: *hyvin* vs. *kovin* $z = 3,46109$ ($p < 0,001$); *kovin* vs. *oikein* $z = 2,268077$ ($p < 0,05$).

⁵⁰ *hyvin* vs. *kovin* $z = 2,25448$ ($p < 0,05$);

⁵¹ Lasken indefiniittiset pronominit *jokin*, *kaikki* ja *moni* sekä adverbis *paljon* ja *vähän* tässä työssä Hakulisen ja Karlssonin (1979: 81–82) ja Vilkun (1996: 52, 214–222) tavoin kvanttoreiksi.

Konkordanssi 6.2. Esimerkkejä konjunktiokollokaateista ja niiden asemasta kotekstissa.

1	an, viisissäkymmenissä ja aina muulloin	<i>hyvinkin</i>	elegantti ja hiukan pidättyväinen
2	un huuto: - Menikö oikein? - Meni. Olin	<i>hyvin</i>	onnellinen, mutta myös syvästi li
3	t täplittivät taivasta. Tiesin, että tähdet.	<i>hyvin</i>	korkealla. Mutta kenties huikeat,
4	tä syövereistä, kirkastui ja muuttui äkkiä	<i>kovin</i>	eläväksi, melkeinpä käsin kosketel
5	uijotteli ikkunasta ulos. Tuntui siltä että	<i>kovin</i>	kauas se ei uskaltanut katsoa. Nen
6	sta. Ja löi krusifiksilla lehtoria otsaan, ei	<i>kovin</i>	lujaa mutta tuntuvasti niin että ot
7	ä elämää. Juoksin ulos ja hengitin tuulta	<i>oikein</i>	syvästi. Tuuli oli minun iso kaveri.
8	myös sanomaan Askolle, että jos rupeaa	<i>oikein</i>	mahdottomalta tuntumaan, niin ka
9	t sen. - Ai jai, taas pääsi rouva pistämään	<i>oikein</i>	pahasti. Mutta totellaan, totellaan

Konkordanssi 6.3. Esimerkkejä astemääritteiden pronominkollokaateista.

1	järveltä alkoi nousta usvaa. -- Minä olin	<i>hyvin</i>	hämmäntynyt sen ensimmäisen ta
2	armasti? Se on vuorenvarmaa. Minä olen	<i>hyvin</i>	voimakas. Minulle ei pärjää kukaa
3	vista velvollisuuksista. Hänen ihonsa oli	<i>hyvin</i>	herkkä. Hän hämmäntyi, kun tapu
4	sä kauniisti. Ei se kuitenkaan ole minulle	<i>kovin</i>	tärkeätä. Olen kaupunkilainen. M
5	annekaan ole värejä, vain muotoa. Se on	<i>kovin</i>	maskuliininen. Hän arveli, että vi
6	mmutettu. Ensimmäinen kohde odottaisi.	<i>Kovin</i>	röyhkeäksi hän ei varmaankaan us
7	Kummisetä ravisti päätään. - Minä voin	<i>oikein</i>	hyvin, hän sanoi pontevasti, - olo
8	tanssijan vartalo. Ja tämän minä muistan	<i>oikein</i>	hyvin. Katson Jelenaa ja sanon it
9	zzaan halua matkustaa, Asko sanoi ja oli	<i>oikein</i>	tyytyväinen itseensä . Taisipa iskeä

Substantiivikollokaatit. Määritteiden *hyvin* ja *kovin* vähintään viisi kertaa esiintyvät substantiivikollokaatit ovat yhtä poikkeusta lukuun ottamatta ihmistarkoitteisia substantiiveja. Näiden kollokaattien osuus kaikista substantiivikollokaateista on kuitenkin varsin pieni (2 % ja 3 %), mikä merkitsee sitä, että ryhmä sisältää runsaasti variaatiota. *oikein*-määritteellä ei sen sijaan ole ainuttakaan vähintään viisi kertaa esiintyvää substantiivikollokaattia kotekstissaan:

hyvin <MIES, IHMINEN, NAINEN 2 %>

kovin <MIES, NAINEN, VUOSI 3 %>

oikein <->

Kollokaatteja on tasapuolisesti sekä vasemman- että oikeapuoleisessa kotekstissa, kuten konkordanssin 6.4 esimerkeistään näkee.

Konkordanssi 6.4. Substantiivikollokaatteja *hyvin*- ja *kovin*-määritteiden kontekstissa.

1	oinen taempänä istuvista miehistä oli	<i>hyvin</i>	kookas, melkeinpä jättiläinen mitt
2	ttelevään lannevaatteeseen. Nainen oli	<i>hyvin</i>	laiha. Kylkiluut pystyi erottama
3	på finska. Minun poikani on aina ollut	<i>hyvin</i>	itsenäinen ihminen , oman tiensä
4	in yleisesti ottaen olen pohjimmiltani	<i>hyvin</i>	viehättävä mies , ja naiset rakast
5	tä, että miesten tiedettiin viihtyvän	<i>kovin</i>	hyvin keskenään eikä olevan kiin
6	putoamaan. Ehkä Regina auttoi. Kaksi	<i>kovin</i>	erilaista naista olivat ystävyyst
7	oret oli häädetty poliisivoimin, mutta	<i>kovinkaan</i>	monta vuotta kaupunki ei ollut u

Kvanttorikollokaatit. Kaikilla kolmella synonyymilla on kontekstissaan kvanttorikollokaatteja, yleisimpien kirjo ja osuudet vaihtelevat astemääritteittäin seuraavasti:

hyvin <JOKIN 30 %, *vähän* 19 %, MONI, *paljon* 12 %>

kovin <*paljon* 38 %, MONI 32 %>

oikein <KAIKKI 28 %>

hyvin ja *kovin* jakavat paljouskvanttorin *paljon* ja lukumääräkvanttorin MONI, joskin näiden suhteelliset osuudet ovat selvästi eri suuret ja erot tilastollisesti merkitseviä⁵². JOKIN ja *vähän* esiintyvät taajaan vain *hyvin*-määritteen ja KAIKKI puolestaan vain *oikein*-määritteen kontekstissa. Näistä JOKIN ja KAIKKI ovat aina vasemmanpuoleisessa kontekstissa (yleisimmin positioissa V1), kun taas *paljon*, MONI ja *vähän* tyypillisimmin positioista O1 (konkordanssi 6.5):

Konkordanssi 6.5. Kvanttorikollokaattien käyttötapaesimerkkejä.

1	ista teitä. Vielä Ension puhuessa jokin	<i>hyvin</i>	kaukainen, muiston tapainen viesti
2	inulle varjon josta en pääse. Ja jossakin	<i>hyvin</i>	syvässä mielessä jota mikään vulga
3	täjä. Hän tiesi Pizzaperheen liiketoimista	<i>hyvin</i>	vähän eikä halunnut tietää enem
4	öityen Paavolan sanat. Viime aikoina oli	<i>kovin</i>	paljon kirjoitettu pankinjohtajista,
5	. Kysyin mistä naisen saa kiinni. Ei niitä	<i>kovin</i>	monta luettelossa ole, tonnimisiä, k
6	oli valmistautunut hyvin. Kaikki olivat	<i>oikein</i>	hyvin valmistautuneet. Siunaus toi

Adverbikollokaatit. Adverbikollokaateista *hyvin*, *nyt* ja *sitten* yhdistävät *hyvin*- ja *oikein*-määritteitä, *niin*-adverbi *hyvin*- ja *kovin*-määritteitä; eniten erilaisia adverbikollokaatteja on *hyvin*-määritteellä, mutta tämä voi jälleen johtua osittain kyseisen määritteen suuremmasta frekvenssistä. *oikein*-määritteen kollokaateista *hyvin* on silmiinpistävästi muita yleisempi: sen osuus on 32 prosenttia (n = 27) kaikista adverbikollokaateista ja peräti 4,5 prosenttia kaikista kollokaateista (taulukko 6.5), eli se on lähes yhtä taajaan esiintyvä kollokaatti *oikein*-määritteen kontekstissa kuin

⁵² *moni*: *hyvin* vs. *kovin* z = 2,22582 (p < 0,05).

paljon: *hyvin* vs. *kovin* z = 2,73942 (p < 0,01).

kielen yleisin sana OLLA. Muut määritteet eivät osoita yhtä selvää preferenssiä johonkin adverbikollokaattiin:

hyvin <*hyvin* 6 %, *hitaasti*, *niin* 4 %, *hiljaa*, *sitten*, *vain* 3 %, *nyt*, *pian*, *varovasti* 2 % >

kovin <*niin* 7 %, *usein* 6 %>

oikein <*hyvin* 32 %, *nyt*, *sitten* 6 %>

Vaikka kollokaatit *niin* ja *hyvin* osittain yhdistävät tutkittavia synonyymeja, ovat niiden käyttötavat erilaisia: Ensinnäkin *niin* on *hyvin*-määritteen kontekstissa varsin monessa funktiossa, kuten määrää tai astetta ilmaisemassa astemääritteenä (konkordanssissa 6.6 rivi 1), liittokonjunktion osana asioiden samantapaisuutta osoittamassa (rivi 2) tai diskurssipartikkelina (rivi 3), kun taas *kovin*-määritteen kontekstissa se on tyypillisimmin modaalisen adverbina (rivi 4). *hyvin* on vuorostaan *hyvin*-määritteen kontekstissa ollessaan aina astemääritteenä toistorakenteessa (rivit 5 ja 6). Tällöin on kyse selvästä astemääritteen itsekollokoinnista (Jones ja Sinclair 1974: 45–47), joka syntyy tekstissä käytetystä retorista toistosta. *oikein*-määritteellä *hyvin* on taas tekemisen tapaa ilmaisevan materiaalsen adverbien funktiossa (rivit 7 ja 8). Adverbit erottavat synonyymit toisistaan siis sekä frekvenssinsä että funktionsa perusteella.

Konkordanssi 6.6. Adverbikollokaattien käyttötapaesimerkkejä.

1	ki ilmestyi näkyviin. Sitten hän kuiskasi	<i>hyvin</i>	hiljaa, niin hiljaa että Lounerin
2	stan viehättävillä kesakoilla. Iho oli siis	<i>hyvin</i>	vaalea niin kuin koko Vahon pe
3	, hän kuiskasi. - joskus ne saattavat olla	<i>hyvinkin</i>	räyhäkkäitä. - Niinkö? Ensio hörä
4	htään mitään, mies naurahti. Se oli niin	<i>kovin</i>	yksinkertaista, viisitoista vuotta si
5	seni takia. Se on hyvin houkuttava halu,	<i>hyvin</i>	voimakas, hyvin kiihottava. Luu
6	stympään asentoon. Sitten hän kääntyi	<i>hyvin</i>	hyvin hitaasti katsomaan Louneria
7	aavat saada pahansuopaa puhetta aikaan	<i>oikein</i>	hyvin , jos vain ahkerasti tahtovat.
8	kä minäkään ehdi. Minä viihdyn kotona	<i>oikein</i>	hyvin . Minä voin erinomaisesti, p

Adjektiivikollokaatit. Adjektiivikollokaatit erottelevat sanat toisistaan edellisiä kollokaattiryhmiäkin selvemmin: astemääritteet eivät jaa ainuttakaan adjektiivia vähintään viidesti esiintyvien kollokaattien luettelossa. Kollokaatit ovat alla.

hyvin <KAUNIS, NUORI, PIENI, VANHA, VÄSYNYT 2 %>

kovin <TÄRKEÄ 4 %>

oikein <HYVÄ 22 %, ISO 5 %>

Kollokaattien osuudet eivät ole kovin suuria, paitsi HYVÄ-adjektiivin, joka selvästi erottuu *oikein*-määritteen tyypillisimmäksi adjektiivikollokaatiksi: se on edellä käsitellyn *hyvin*-adverbien tapaan muita selvästi yleisempi (22 %, n = 21) ja kaikista kollokaateista kolmanneksi yleisin 3,5 prosentin osuudellaan (taulukko 6.5). Adjek-

tiivikollokaateista esiintyvät vain paria poikkeusta lukuun ottamatta kaikki positiossa O1, jolloin astemääritteet määrittävät adjektiivin ilmoittavan ominaisuuden astetta (konkordanssi 6.7).

Konkordanssi 6.7. Esimerkkejä frekventeistä adjektiivikollokaateista.

1	ti ovi toiseen huoneeseen, jossa oli vain	<i>hyvin</i>	pieni ikkuna ja korkealle seinälle k
2	n pienen asiakaskuntansa keskuudessa	<i>hyvin</i>	suosittu. Tähän vaikutti mämmin p
3	kit kuitenkin ovat. Tuntuu kuin katselisi	<i>hyvin</i>	vanhan ihmisen kasvoja. Huonosta
4	lloa. Menimme uudestaan sänkyyn. Olin	<i>hyvin</i>	väsynyt . Sari oli kainalossani, mu
5	staan. - Sie oot saanu päähän tällin. - Jo	<i>hyvin</i>	nuorena . - Mie soitan sairausauton. -
6	et kaunis, hän sanoi. - Juuri nyt sinä olet	<i>hyvin</i>	kaunis . - Älä valehtele, Regina to
7	hmentäen, - että tämä paikka on minulle	<i>kovin</i>	tärkeä ja rakas juuri tällaisena, ei
8	ä kauniisti. Ei se kuitenkaan ole minulle	<i>kovin</i>	tärkeätä . Olen kaupunkilainen. M
9	tehit iltapäivälä vielä toisen reissun. Jos	<i>oikein</i>	hyvä marjapounukko oli kohalle sat
10	Raikuvia kättentaputuksia auditoriossa.	<i>oikein</i>	hyvää matkanjatkoa kaikille. Käytt

Yllä oleva tarkastelu paljastaa, että synonyymisilla astemääritteillä on selvää toisistaan poikkeavaa kollokationaalista preferenssiä. Tämä tuli ilmi jossain määrin kaikissa tarkastelluissa kollokaattiryhmissä kollokaattien yleisyysjärjestyksessä ja suhteellisissa osuuksissa tai funktioissa, joissa kollokaatteja oli käytetty. Tyypillisimmin kontekstuaalinen valinta osoittautuu nimenomaan *preferenssiksi*, ei rajoitukseksi (ks. lukua 2.3.1). Muutamassa tapauksessa voidaan kuitenkin nähdä rajoituksen kaltaista kollokationaalista valintaa. Tällainen on ensinnäkin *hyvin-* ja *oikein-*määritteiden erittäin voimakas negatiivinen preferenssi kieltöhakuisuuteen, mikä ilmenee siten, että ne kollokoivat erittäin harvoin (0,6–2,5 %) kieltöverbin kanssa samassa lauseessa (taulukko 6.6). Muutamat kieltöverbillisistä lauseista tavattavat astemääritteet osoittavat, että kyse ei aivan ole rikkomattomasta rajoituksesta, mutta näin tuskin on koskaan minkään säännön ja rajoituksen kohdalla, vaan poikkeukset ovat jossain määrin aina mahdollisia – etenkin aineistoni kaltaisissa kaunokirjallissa teksteissä. *kovin-*astemääritteellä kieltö- ja myöntöhakuisuuden preferointisuhteet eivät ole yhtä jyrkkiä, vaan määritettä käytetään sekä myöntö- että kieltolauseissa, joskin esiintyminen kieltolauseissa on selvästi tavallisempaa, mikä kielii voimakkaasta preferenssistä.

Selkeä kollokationaalinen rajoitus liittyy *hyvin-*kollokaattiin: kun *hyvin-*määrite saa kontekstiinsa *hyvin-*adverbin pelkästään toistorakenteissa olevana astemääritteenä, niin *oikein-*määritteen kontekstissa *hyvin* on aina materiaalisena adverbina. Tämä johtuu pääasiassa siitä, että määritettäessä tapaa ilmaisevaa *hyvin-*adverbia (‘mainiosti, erinomaisella tavalla’) ominaisuuden suurta määrää ilmaisevaksi vahventajaksi on vakiintunut (näistä kolmesta määritteestä) *oikein*, jolloin vältetään tautologinen määrittäminen **hyvin hyvin. kovinkaan* ei tähän tunnu kelpaavan kieltöhakuisuutensa vuoksi, ellei sitten kielletä tekemisen tavan mainioutta (*Teillä ei taida mennä kovin hyvin*). Kolmas rajoitus koskee adjektiivin *HYVÄ* määrittämistä. Tähänkin määrittämissuhteeseen on vakiintunut *oikein*, luultavimmin tautologisen *hyvin hyvä* -ilmauksen karttamisen vuoksi. Kyse on siis jälleen pikemminkin rajoituksesta

kuin preferenssistä – toki toisteisetkin myötäesiintymät ovat mahdollisia (kollokaatio *hyvin hyvä* esiintyy aineistossa kerran), mutta silloin kyse on selvästi poikkeuksesta.

Lopuksi olen koonnut astemääritteiden synonyymisyyttä korostavat kollokaatit eli sellaiset myötäesiintymät, jotka ovat jokaisen tässä käsitellyn astemääritteen kontekstissa, positioittain taulukkoon 6.7. Synonyymisyyttä korostavien kollokaatioiden määrä vähenee selvästi sitä mukaa, mitä lähemmäksi astemääritettä saavutaan, ja tilanne on samanlainen noodioiden molemmilla puolilla. Näin ollen synonyymisyyttä rikkovia kollokaatteja tavataan useammin astemääritteiden vieruskollokaateissa ja synonymiaa korostavia puolestaan tarkastelualueen reunoilla. Tämän perusteella voisi olettaa, että lekseemille ominaiset kontekstuaaliset leksikaaliset piirteet, jotka erottavat ne merkitykseltään läheisistä sanoista, löytyvät tyypillisimmin juuri aivan lähikontekstista, jopa vieruspositioista. Tämä ei luonnollisestikaan ole ristiriidassa sen tosiasian kanssa, että kyseisiä kollokaatteja tavataan myös lähipositioita kauempana.

Taulukko 6.7. Astemääritteitä yhdistävät kollokaatit positioittain.

V3	V2	V1	noodi	O1	O2
olla	olla	olla		paljon	ja
ja	minä	minä			mutta
että	se				minä
minä					hän
se					se
					ei

6.1.2. Kollokationaalinen diversiteetti

Tekstien leksikaalista diversiteettiä tutkitaan usein *type–token*-suhteen (*type–token ratio*, TTR) avulla. Suhde kuvaa tekstien sanastollista kertautuvuutta (ks. mm. Barnbrook 1996: 53–54; Stubbs 2001a: 133–134), ja siinä verrataan tekstissä esiintyvien eri sananmuotojen lukumäärää koko tekstin sanemäärään. Suhdetta on käytetty viime aikoina kuvaamassa muun muassa käännosten ja alkuperäistekstien (Zanettin 2000), käännoisenglannin ja alkuperäisenglannin (Laviosa 1996) sekä suomen osalta oppikirjatekstien ja tiedetekstien (Karvonen 1995) leksikaalista diversiteettiä ja myös optimaalisen tarkastelualueen kokoa (Mason 1997, 2000; ks. lukua 2.3.2). Sanastollista rikkautta on kuitenkin suomen kielen kaltaisissa synteettisissä, monia taivutusmuotoja sisältävässä kielissä parempi laskea lemma–sanasuhteen (LSS) avulla, jossa lasketaan tekstissä olevien lemموjen (lekseemien) suhde koko tekstin sanemäärästä. Tällöin kuvataan nimenomaan sanastollista rikkautta, ei sananmuotojen rikkautta. LSS:n laskeminen on ollut tekstien kuvauksessa

TTR:ää harvinaisempaa, koska tutkijoiden käytössä olleet aineistot ovat harvoin olleet lemmatisoituja (Granger ja Wynne 2000: 251).

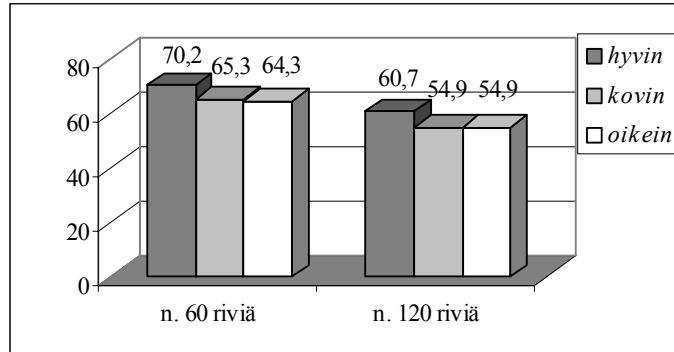
Seuraavassa analyysissä lemma–sane-suhdetta sovelletaan synonyymien kotekstin leksikaalisen varianssin tutkimiseen siten, että a) koko tarkastelualueen V3–O2 eri lemموjen määrää on verrattu alueen koko sanemäärään ja b) yhden position eri lemموjen määrää on verrattu kyseisen position koko sanemäärään. LLS (myös TTR) on sitä pienempi, mitä pitempi teksti on. Tämä johtuu siitä, että vaikka saneiden määrä kasvaa tekstin pidetessä, lemموjen (ja sananmuotojen) määrä ei kasva samassa suhteessa, vaan ne toistuvat tekstissä (Stubbs 2001a: 133). Tästä syystä vertailtavien tekstimassojen olisi oltava suunnilleen yhtä pitkiä.

Omassa tutkimuksessani astemääritteitä ympäröivän tekstimassan määrä vaihtelee suuresti sen mukaan, miten frekventti itse määrite on: *hyvin*-määritteen tarkastelualueen koteksti on 1760 sanetta, mutta *oikein*-määritteen vain 605 sanetta, joten näiden alueiden leksikaalisen diversiteetin vertailu ei olisi mielekäästä sellaisenaan, koska on oletettavissa, että sanastollinen variaatio riippuisi tässäkin tekstimassan määrästä. Tästä syystä tarkasteluun otettavaa tekstimassaa on tasoitettu samankokoiseksi. Diversiteetin tarkastelua varten on tehty kaksi diversiteetin testausta: toista varten on laskettu mukaan noin 60 ja toista noin 120 astemääritettä koteksteineen, jolloin tarkasteltava sanemassa on ensimmäisessä testissä noin 300 sanetta ja toisessa noin 600 sanetta. Kahta testausta käyttämällä pyritään hakemaan tulosten luotettavuutta, koska aineisto on suhteellisen pieni; toisaalta näin menetellen voidaan myös etsiä sopivaa tekstimassan määrää diversiteetin mittaukseen. Ensimmäistä testiä varten on *hyvin*-määritteen esiintymistä valittu konkordanssista joka kuudes rivi, *kovin*-määritteeltä joka kolmas ja *oikein*-määritteeltä joka toinen rivi, jolloin konkordanssirivien määräksi tuli noin 60. Toista testiä varten valittiin *hyvin*-määritteen konkordanssista joka kolmas rivi, *kovin*-määritteeltä vuorostaan poistettiin joka kolmas rivi ja *oikein*-määritteeltä otettiin kaikki korpuksen esiintymät mukaan, jolloin päädyttiin noin 120 esiintymään. Tulokset on esitetty kuviossa 6.1 (s. 100). Suhdeluku on laskettu siten, että yksi lemma on laskettu vain kerran ja kaikkien lemموjen määrä on jaettu konkordanssirivien kotekstin saneiden määrällä. Tämä on kerrottu vielä sadalla, joten käytännössä suhdeluku kertoo lemموjen prosentuaalisen osuuden.

Molemmat testit antavat samanlaisen tuloksen: *hyvin*-määritteen kotekstin sanasto näyttää olevan varioivampaa kuin sen synonyymien, joiden arvot ovat keskenään lähes samat. Tilastollisen testauksen mukaan 60 rivin otoksesta tehdyssä testauksessa ilmenevät erot eivät ole merkitseviä, mikä saattaa johtua materiaalin vähyydestä. 120 rivin otoksessa näkyvät erot parien *hyvin*–*kovin* ja *hyvin*–*oikein* välillä ovat tilastollisesti merkitseviä⁵³. Koska ensimmäisenkin tekstin suuntaus on samanlainen, voidaan pitää osoitettuna, että *hyvin*-määritteen kollokaattien varianssi on monipuolisempaa kuin sen kahden synonyymien, joiden välillä ei taas ole eroa.

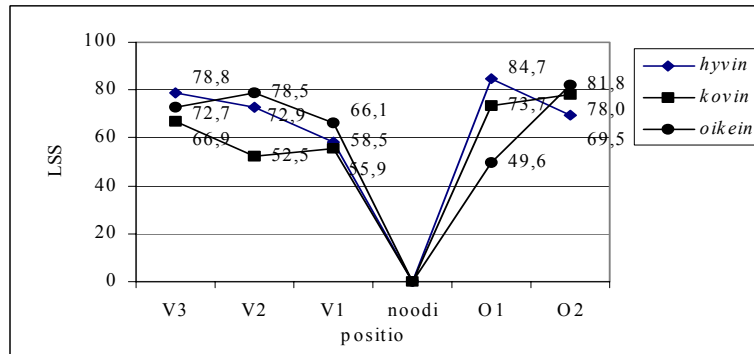
⁵³ n. 60 riviä: *hyvin* vs. *kovin* $z = 1,261617$ ($p > 0,05$); *hyvin* vs. *oikein* $z = 1,45013$ ($p > 0,05$); *kovin* vs. *oikein* $z = 0,275778$ ($p > 0,05$).

n. 120 riviä: *hyvin* vs. *kovin* $z = 2,004071$ ($p < 0,05$); *hyvin* vs. *oikein* $z = 2,029972$ ($p < 0,05$); *kovin* vs. *oikein* $z = 0,013623$ ($p > 0,05$).



Kuvio 6.1. Lemma-sane-suhde tarkastelualueella V3-O2.

Samalla menetelmällä voidaan tarkastella myös positiokohtaista variaatiota. Aiemmin tuli ilmi, että määritteillä oli joissakin asemissa kollokaatteja, jotka olivat varsin taajaan esiintyviä. Tällaisia olivat muun muassa *hyvin* ja *HYVÄ oikein*-määritteen positiossa O1. Tällaisten suurifrekvenssisten kollokaattien esiintymisestä voi olettaa, että kollokaattien kirjo vaihtelee astemääritteillä myös asemittain. Tämän tarkistamiseksi on tehty sanastollisen diversiteetin testaus positioittain noin 120 konkordanssirivin tekstimassasta. Tulokset näkyvät kuviosta 6.2.



Kuvio 6.2. Lemma-sane-suhteet positioittain.

Astemääritteiden kollokationaalinen diversiteetti osoittautuu yllättävänkin erilaiseksi positioittain tarkasteltuna. Ensinnäkin astemääritteet eroavat sen suhteen, missä positioissa niillä on vähiten ja missä eniten sanastollista varianssia. *oikein*- ja *kovin*-määritteillä on eniten variaatiota positioissa O2 (LLS 81,8 ja 78,0), *hyvin*-määritteellä vastaavasti asemassa O1 (84,7). *oikein*-määritteellä on puolestaan vähiten kollokationaalista kirjoa oikeanpuoleisissa vieruskollokaateissa (O1: 49,6), kun taas *kovin*- ja *hyvin*-määritteillä vasemmanpuoleisissa positioissa (*hyvin* V1: 58,5 ja *kovin* V2: 52,5). Kokonaisuutena katsottuna *kovin* poikkeaa synonyymeista selvästi sanastollisen varianssinsa suhteen: sen kaikissa vasemmanpuoleisissa positioissa arvot ovat synonyymien arvoja pienemmät. Tämä kertoo, että sen va-

semmanpuoleinen kollokationaalinen diversiteetti on kaikkinsa pienempi kuin sen synonyymeilla. Eniten synonyymeja erottavat positiot V2 ja O1, muissa asemassa erot eivät ole yhtä selvät.

Positiokohtainen kollokationaalinen diversiteetti on sidoksissa taajimmin esiintyvien kollokaattien määrään. Esimerkiksi vain *kovin*-määritteellä on selvästi kaksi muita frekventimpää kollokaattia (EI ja OLLA) positiossa V2, mikä näkyy edellä olevassa kuviossa 6.2 siinä, että V2-position kollokationaalinen kirjo on *kovin*-määritteellä muita selvästi pienempi. Vastaavasti taas *oikein*-määritteen oikeanpuoleisten vieruskollokaattien pienen kollokaatiovarianssin selitys löytyy osittain siitä, että sekä adverbi *hyvin* että adjektiivi HYVÄ ovat erittäin taajaan esiintyviä kollokaatteja, mikä vähentää muiden kollokaattien osuutta koko kollokaattikirjossa. Taulukon 6.4 mukaan *kovin*-määritteen kollokaatit EI ja OLLA sekä *oikein*-määritteen *hyvin* ja HYVÄ ovat myös tilastollisesti merkitseviä kollokaatteja. Asetelma näyttää viittaavaan siihen, että *kovin*-määritteen kollokationaalinen sidoksisuus on vahvaa erityisesti vasemmalle samalla kun sen kollokationaalinen diversiteetti on vasemmalla vähäisintä, ja vuorostaan *oikein* on – päinvastoin kuin sen synonyymit – kollokationaalisesti sidoksissa eniten oikeanpuoleisiin vieruskollokaatteihinsa ja samalla sen kollokationaalinen kirjo on vähäisintä O1-asemassa (vaikkakin jälleen suuri O2-asemassa). *hyvin* näyttää asettuvan näiden kahden ääripään välille, sillä sillä on kiinteimmät kollokationaaliset sidokset positioon V1, jossa muidenkin astemääritteiden kollokationaalinen diversiteetti on suhteellisen vähäistä. Suuri varianssi positiossa O1 taas kertoo siitä, että *hyvin*-määritettä ei käytetä vain tiettyjen leksiemien määritteenä, vaan sen määrittämissuhteet ovat suhteellisen vapaita – erityisesti *oikein*-määritteeseen verrattuna. Positiokohtainen sanastollisen diversiteetin tarkastelu luonnehtii astemääritteitä entistä selvemmin ja paljastaa synonyymien positiokohtaisen kollokationaalisen diversiteetin erot synonyymeja erottavaksi tekijäksi.

6.1.3. Synonyymiset *hyvin*, *kovin* ja *oikein* klustereissa

Viimeinen kollokaatteihin liittyvä tarkastelu koskee astemääritteiden kontekstissa ilmenevää klusteroitumista. Seuraavaan esitykseen on koottu kolmen astemääritteen konkordansseista kaikki vähintään kahden eri tekstintuottajan käyttämät klusterit. Ne on koottu astemääritteittäin taulukoihin 6.8, 6.9 ja 6.10. Kussakin taulukossa on klusterin frekvenssi ilmoitettu oikeassa sarakkeessa. Taulukko 6.8 kuvaa *hyvin*-määritteen klusteroitumisen. Nelisanaisia klustereita on tällä määritteellä vain yksi, kolmisania taas 25 erilaista. Tyypillisin *hyvin*-määritteen klusterirakenne on <SE OLLA *hyvin*>. Lähes yhtä yleisiä ovat myös rakenteet, joissa SE-pronominin paikalla on HÄN tai MINÄ. Koska V2-asemassa on myös muita pronomineja, selvenee *hyvin*-määritteen tyypilliseksi klusterirakenteeksi <pronominikollokaatti OLLA *hyvin*>. Yleinen on myös rakenne <OLLA *hyvin* adjektiivikollokaatti>. Positio V1 on selvästi OLLA-lemman asema: peräti 17 eri klusterissa on OLLA tässä asemassa.

Taulukko 6.8. *hyvin*-määritteen klusteroituminen.

	sanoja	V3	V2	V1	noodi	O1	O2	f
1	4	että	minä	olla	<i>hyvin</i>			2
1	3		se	olla	<i>hyvin</i>			13
2			hän	olla	<i>hyvin</i>			11
3			minä	olla	<i>hyvin</i>			11
4				olla	<i>hyvin</i>	väsytynyt		5
5			tämä	olla	<i>hyvin</i>			4
6			katsoa	minä	<i>hyvin</i>			4
7			olla	jokin	<i>hyvin</i>			4
8			olla	olla	<i>hyvin</i>			3
9				olla	<i>hyvin</i>	kaunis		3
10					<i>hyvin</i>	kauan	sitten	3
11			sinä	olla	<i>hyvin</i>			2
12			he	olla	<i>hyvin</i>			2
13			saattaa	olla	<i>hyvin</i>			2
14			olla	silloin	<i>hyvin</i>			2
15				olla	<i>hyvin</i>	hiljainen		2
16				olla	<i>hyvin</i>	itsenäinen		2
17				olla	<i>hyvin</i>	kiinnostunut		2
18				olla	<i>hyvin</i>	kuuma		2
19				olla	<i>hyvin</i>	kylmä		2
20				olla	<i>hyvin</i>	laiha		2
21				tulla	<i>hyvin</i>	lähellä		2
22				vain	<i>hyvin</i>	harva		2
23				ja	<i>hyvin</i>	pian		2
24					<i>hyvin</i>	kaukainen	ilta	2
25					<i>hyvin</i>	aikaisin	aamu	2

O1-asemassa on tyypillisimmin adjektiivi; positioon O2 ulottuvien klustereiden määrä on taas niin vähäinen, että asemasta ei voi tehdä yleistyksiä. Kaiken kaikkiaan tyypillisin rakenne, eli pronominin, OLLA-kollokaatin ja *hyvin*-astemääritteen muodostama kombinaatio sijoittuu alueelle V2–hakusana.

Kun halutaan tietää, miten monessa konkordanssirivissä eli 352 esiintymässä *hyvin* on osa klusteria, on rypäät etsittävä konkordanssista. Tulokseksi saadaan luku 81, eli *hyvin* esiintyy kahta sanaa pitemmissä, toistuvissa sanarypäissä 23,0 prosentissa kaikista tapauksista, jolloin lähes joka neljännessä esiintymässään se on osa toistuvaa klusteria.

kovin-määritteen sanarypäät on kerätty vuorostaan taulukkoon 6.9. Sen klusteroituminen poikkeaa edellä käsitellystä määritteestä usealla tapaa. Ensinnäkin sillä on peräti viisi erilaista nelisanaista klusteria, kun *hyvin*-määritteellä on vain yksi. Kun vielä otetaan huomioon, että *kovin*-määritteen frekvenssi ($n = 176$) on puolta pienempi kuin *hyvin*-määritteen ($n = 352$), kasvaa nelisanaisten klustereiden merkitys: ne osoittavat, että *kovin* on taipuvainen esiintymään synonyymiaan selvästi useammin pitemmissä sanaryppäissä. *kovin*-määritteen tavallisimmat klusterit sijoittuvat myös alueelle V2–noodi, kaikkiaan lähes puolet kaikista kolmisanaista

Taulukko 6.9. *kovin*-määritteen klusteroituminen.

	sanoja	V3	V2	V1	noodi	O1	O2	f
1	4	hän	ei	olla	<i>kovin</i>			3
2		minä	ei	olla	<i>kovin</i>			2
3		ei	se	olla	<i>kovin</i>			2
4		olla	viime	aika	<i>kovin</i>			2
5			olla	minä	<i>kovin</i>	tärkeä		2
1	3		ei	olla	<i>kovin</i>			24
2			se	olla	<i>kovin</i>			7
3				olla	<i>kovin</i>	moni		4
4				olla	<i>kovin</i>	paljon		4
5			ei	se	<i>kovin</i>			3
6			ei	kyllä	<i>kovin</i>			3
7			mutta	ei	<i>kovin</i>			3
8					<i>kovin</i>	moni	vuosi	3
9			että	se	<i>kovin</i>			2
10			ei	voida	<i>kovin</i>			2
11			ei	näyttää	<i>kovin</i>			2
12			olla	niin	<i>kovin</i>			2
13			olla	edes	<i>kovin</i>			2
14			olla	minä	<i>kovin</i>			2
15			viime	aika	<i>kovin</i>			2
16				olla	<i>kovin</i>	helppo		2
17				olla	<i>kovin</i>	innostunut		2
18			ei		<i>kovin</i>	paljon		2
19				minä	<i>kovin</i>	tärkeä		2
20					<i>kovin</i>	paljon	kiinnostaa	2
21					<i>kovin</i>	tärkeä	ja	2

klustereista on tällä alueella. *kovin* näyttää olevan kuitenkin vielä enemmän vasemmalle klusterioiva kuin *hyvin*-määrite, mitä korostavat myös nelisanaiset klusterit.

Entäpä klustereiden sanasto? Tyypillisin klusteri on kolmisanaa <EI OLLA *kovin*>. Positiossa V2 on tyypillisimmin EI ja siitä oikealle kohti hakusanaa OLLA. Positiossa O1 on *hyvin*-määritteen tapaan adjektiiveja, mutta erityisesti kvanttoireita. MONI ja *paljon* ovatkin tavallisimmat tämän position kollokaatit kolmisanaissa sanarypäissä, ja klusteri <OLLA *kovin* kvanttorikollokaatti> on toiseksi yleisin klusterirakenne. Positioon O2 yltävät klusterit ovat tälläkin astemääritteellä harvinaisia.

Kun taulukosta palataan konkreettiseen konkordanssiin, voidaan jälleen laskea hakusanan esiintyminen klustereissa koko tekstimassassa. Nyt tuloksena on 68 konkordanssiriviä, mikä tarkoittaa, että *kovin*-määritteen esiintymistä peräti 38,6 prosenttia on sellaisia, joissa se tavataan osana klusteria. Luku on selvästi suurempi kun *hyvin*-määritteellä, sillä yli joka kolmas esiintymistä on klusteri.

Viimeiseksi tarkastellaan määritteen *oikein* klusteroitumista, joka on kuvattu taulukossa 6.10. Myös *oikein* poikkeaa klusteroitumiseltaan muista. Sanarypäiden sanasto on ensinnäkin toisenlainen ja myös vähemmän varioiva. Asemassa V2 on

Taulukko 6.10. *oikein*-astemäärittteen klusteroituminen.

	sanoja	V3	V2	V1	noodi	O1	O2	f
1	4		se	tuntua	<i>oikein</i>	hyvä		2
2			se	olla	<i>oikein</i>	hyvä		2
1	3			olla	<i>oikein</i>	hyvä		5
2			se	olla	<i>oikein</i>			3
3				voida	<i>oikein</i>	hyvin		3
4				muistaa	<i>oikein</i>	hyvin		3
5			se	tuntua	<i>oikein</i>			2
6			tämä	olla	<i>oikein</i>			2
7				tuntua	<i>oikein</i>	hyvä		2
8				ymmärtää	<i>oikein</i>	hyvin		2
9				kyllä	<i>oikein</i>	hyvin		2
10					<i>oikein</i>	hyvin	hän	2

hyvin-määrittteen tapaan pronomini, tyypillisimmin SE. Positio V1 on taas selvä verbiasema, mutta ei yksistään OLLA-verbin miehittämä, sillä asemasta tavataan myös kollokaatit TUNTUA, VOIDA, MUISTAA ja YMMÄRTÄÄ. Selvästi erilaista synonyymeihin verrattuna on se, että positio O1 on ”varattu” vain kollokaateille *hyvin* ja HYVÄ. Klusterirakenne muodostuukin seuraavanlaiseksi: <(SE) verbikollokaatti *oikein* HYVÄ/*hyvin*>. Synonyymeistaan poiketen *oikein*-määrittteen klusterit eivät ole vasemmanpuoleiseen kollokaatiokenttään pakkautuvia, vaan ne sijoittuvat yleensä alueelle V1–O1. Erilaiset kolmisanaiset klusterit vastaavat vain 14,9 prosenttia (n = 18) kaikista tämän määrittteen esiintymistä, mikä on selvästi vähemmän kuin edellisissä tapauksissa. Nelisanaisia klustereita tällä määritteellä on vain kaksi, mutta suhteutettuna pieneen esiintymämäärään tämä on enemmän kuin *hyvin*-määritteellä, jolla nelisanaisia klustereita oli vain yksi.

Klusterianalyysi paljasti ensinnäkin sen, että *hyvin*- ja *kovin*-määritteille on tavallista esiintyä klustereissa; niiden määritteiden osuus, jotka ovat osa jotakin klusteria, oli suurin *kovin*-määritteellä ja selvästi pienin *oikein*-määritteellä. Määritteistä *hyvin* ja erityisesti *kovin* osoittautuivat vasemmalle klusteroiviksi, *oikein*-määrittteen klusteroituminen oli puolestaan enemmänkin keskittynyt määrittteen molemmin puolin. Klusteroitumisominaisuudet ovat yhdenmukaisia positiokohtaista diversiteettiä kuvaavan kuvion 6.2 (s. 100) kanssa: *kovin*- sekä erityisesti *hyvin*-määritteellä positiokohtainen diversiteetti on suuri noodin oikealla puolella ja pieni vasemmalla puolella. Tämä on yhtenevä tulos klusteroitumissuunnan kanssa, sillä klusteroituminen vasemmalle vakiintuneisiin rakenteisiin heijastaa ainakin osittain leksikaalisen variaation vähäisyyttä vasemmalla. Vastaavasti *oikein*-määrittteen diversiteetti oli pienimmillään O1-positiossa, ja sen klusteroituminen osoittautui keskittyneeksi, kun O1-position taajaan tavatut kollokaatit *hyvin* ja HYVÄ sisältyivät vakiintuneisiin toistuviin rakenteisiin.

Synonyymien nelisanaiset klusterit ovat melko harvinaisia. Erityisesti tämä koskee *hyvin*-määritettä, sillä kahdella sen synonyymilla nelisanaisia klustereita on hieman enemmän. Koska klustereiden merkittävyys on toistuvuuden lisäksi suh-

teessa niiden pituuteen (Kjellmer 1984: 167), on todettava, että *kovin-* ja *oikein-*määritteiden nelisanaiset klusterit antavat jonkinlaisen viitteen siitä, että näiden klusteroitumisherkkyys olisi suurempi kuin niiden synonyymilla. Aineiston pienuus kuitenkin rajoittaa tämän tutkimista, ja nelisanaisten klustereiden merkitys jää tässä analyysissä vähäiseksi, joskin huomionarvoiseksi tulokseksi.

6.1.4. Synonyymien *hyvin*, *kovin* ja *oikein* kollokationaalinen profiili

Edellä käsitellyt synonyymisten astemääritteiden *hyvin*, *kovin* ja *oikein* kollokationaaliset preferenssit ja rajoitukset on koottu taulukkoon 6.11, joka kuvaa astemääritteiden kollokationaalisen profiilin pääpiirteittäin. Astemääritteiden muodostamien kollokaatioiden tarkastelu osoitti, että määritteet eroavat toisistaan niin tilastollisesti merkitsevien kuin frekventtienkin kollokaattien perusteella. Mielenkiintoiseksi nousi kollokationaalisen diversiteetin, erityisesti positiokohtaisen variaation, merkitys synonyymiaa rikkovana tekijänä. Näiden lisäksi astemääritteet osoittavat selvää taipumusta erilaiseen ja eriaistaiseen klusteroitumiseen.

Taulukko 6.11. Astemääritteiden *hyvin*, *kovin* ja *oikein* kollokationaalinen profiili.

	<i>hyvin</i>	<i>kovin</i>	<i>oikein</i>
Kollokaatit			
- frekventein V3–O2	OLLA (9,0 %)	OLLA (10,2 %)	OLLA (4,8 %)
- tilastollisesti merkitsevin (<i>MI</i> -testi / <i>t</i> -testi)	VÄSYNYT / OLLA	<i>usein</i> / EI	<i>hyvin</i> / <i>hyvin</i>
- kollokationaalinen rajoitus	- neg. kollokointi EI (myöntöhakuinen) - neg. kollokointi < <i>hyvin</i> , HYVÄ>		- neg. kollokointi EI (myöntöhakuinen) - < <i>hyvin</i> , HYVÄ>
Kollokationaalisen diversiteetti*			
lemma–sane-suhde	60,7	54,9	54,9
- suurimmillaan positiossa	O1	O2	O2
- pienimmillään positiossa	V1	V2	O1
Klusterit			
- tyypillisin klusteroitumisen suunta	<SE OLLA <i>hyvin</i> > vasemmalle	<EI OLLA <i>kovin</i> > vasemmalle	<OLLA <i>oikein</i> HYVÄ> keskittynyt
- klusteroitumisaste	23,0	38,6	14,9

*N. 120 konkordanssirivin perusteella.

6.2. Kolligationaalinen valinta

Tässä luvussa lähestytään tutkittavien astemääritteiden synonyymisyyttä tarkastelemalla niiden syntagmaattisia suhteita kontekstissa olevien sanaluokkien kanssa ja suhdetta lausepositioon. Jälkimmäinen tarkoittaa sitä, että mikäli astemääritteen jommallakummalla puolella on lauseen alku tai loppu, lasketaan tämä omaksi kolligaatikseen. Sanaluokkien ja lausepositioiden osuuksia tarkastellaan tilastollisen analyysin avulla. Kolligationaalisia suhteita lähestytään jälleen ensin tarkastelemalla koko havaintoaluetta, minkä jälkeen huomio kohdistetaan vasemman- ja oikeanpuoleisiin konteksteihin. Kuten kollokaattejakin tutkittaessa, myös tässä yhteydessä selvitetään astemääritteiden klusteroituminen tietyytyyppeihin kolligaattiklustereihin. Vaikka klusteri-käsitettä onkin käytetty pääasiassa leksikaalisia ketjuja analysoitaessa, käytän sitä tässä yhteydessä myös kolligaattianalyyseissä, koska tarkastelukohteena ovat samaan tapaan toistuvat kahta yksikköä pitemmät kiinteät ketjut. Tarkastelussa keskitytään erityisesti synonyymeja erottaviin tekijöihin ja pyritään löytämään syyt, jotka aiheuttavat eroja synonyymien kolligationaaliseen syntagmaan.

6.2.1. Kolligationaalinen valinta tarkastelualueella

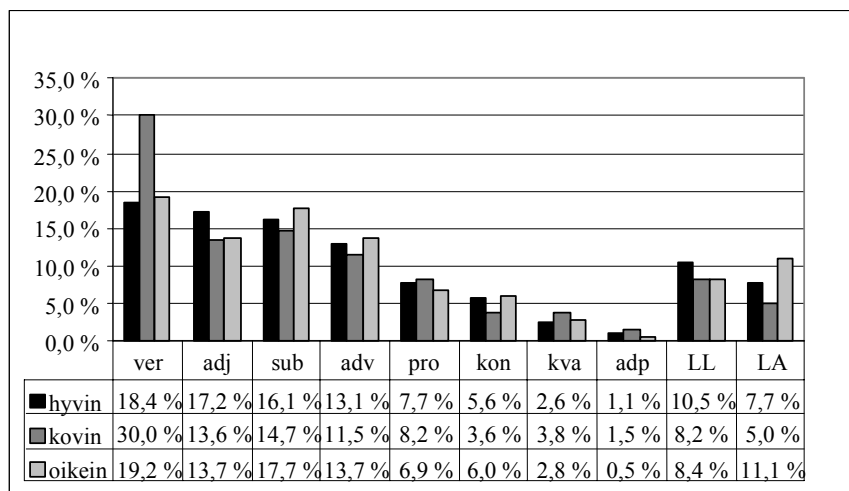
Koko tarkastelualuetta koskeva kolligationaalinen valinta on kuvattu kuviossa 6.3. Sanaluokkakolligaatit näyttävät jakautuvan astemääritteittäin suhteellisen tasaisesti lukuun ottamatta verbien ja lauseenalkuisuuden (LA) määrää, jotka erottuvat jo silmämääräisestäkin erilaisiksi. Ensinnäkin verbien osuus on selvästi suurempi *kovin*-määritteellä kuin sen synonyymeilla: *kovin*-määritteen verbikolligaattien osuus kaikista kolligaateista on 30,0 % (n = 264), kun *hyvin*-määritteellä osuus on 18,4 % (n = 323) ja *oikein*-määritteellä 19,2 % (n = 116). Erot *kovin*-määritteen osuudessa suhteessa muihin ovat tilastollisesti erittäin merkitseviä⁵⁴. Eron aiheuttajaksi voi jo tässä vaiheessa ounastella *kovin*-määritteen frekventtiä kollokaattia kieltoverbiä EI, joka esiintyy sen kontekstissa verbien liittomuodoissa. Lauseenalkuisuus erottaa määritteet siten, että *oikein*-määritteen kolligaateista se muodostaa 11,1 % (n = 67) kun vastaavat osuudet ovat *hyvin*-määritteellä 7,7 % (n = 136) ja *kovin*-määritteellä vain 5,0 % (n = 44). Lauseenalkuisuuskolligaattien osuuksien ero on tilastollisesti merkitsevä kaikkien astemääriteparien välillä, eli se, miten usein lauseenalkuisuus esiintyy määritteiden kontekstissa, erottelee kaikki määritteet toisistaan. Merkitsevin osuuksien ero on *kovin*- ja *oikein*-määritteiden välillä⁵⁵. Tilastollinen testaus paljastaa erottaviksi tekijöiksi vielä konjunktioiden ja adjektiivien osuudet. Näistä ensimmäisten osuus on *kovin*-määritteellä synonyymejaan merkitsevästi pienempi (3,6

⁵⁴ *hyvin* vs. *kovin* z = 6,784648 (p < 0,001); *kovin* vs. *oikein* z = 4,697808 (p < 0,001).

⁵⁵ *hyvin* vs. *kovin* z = 2,620743 (p < 0,01); *hyvin* vs. *oikein* z = 2,535382 (p < 0,05); *kovin* vs. *oikein* z = 4,373502 (p < 0,001).

%, n = 32; *hyvin*-määritteellä 5,6 %, n = 99 ja *oikein*-määritteellä 6,0 %, n = 36)⁵⁶. Adjektiivien osuus erottelee vuorostaan *hyvin*-määritteen synonyymeistaan: tällä adjektiivien osuus on 17,2 % (n = 303), *kovin*- ja *oikein*-määritteillä osuudet ovat pienemmät ja keskenään samansuuruiset, 13,6 % (n = 120) ja 13,7 % (n = 83)⁵⁷. Vaikka muidenkin kolligaattien osuudet poikkeavat toisistaan jonkin verran, niiden erot eivät osoittaudu kuitenkaan tilastollisesti merkitseviksi.

Synonyymisten astemääritteiden koko tarkastelualueella analysoidessa erityisesti *kovin* näyttää erottuvan kolligaattien osuuksien perusteella kahdesta muusta: *hyvin*-määritteestä se poikkeaa neljän kolligaatin osuuksillaan ja *oikein*-määritteestä kolmen kolligaatin osuuksilla. *hyvin*- ja *oikein*-määritettä erottavat toisistaan vain kahden kolligaatin osuudet. Kolligaattien yleisyysjärjestyksessä ei ole radikaaleja eroja astemääritteiden kesken, kun tarkastelualueella katsotaan kokonaisuutena. Tässä esitettyjä kaikkia kolligaatteja tavataan kaikilla määritteillä, ja yleisin niistä on kaikilla verbikolligaatti. Sijalla kaksi ja kolme vaihtelevat substantiivit ja adjektiivit, neljäntenä ovat adverbit. Seuraaviksi yltävät lausepositiokolligaatit (joista kuitenkin lauseenalkuisuus on sijalla viisi *oikein*-määritteen yleisyyslistalla kun muilla sijoitus on seitsemäs) ja pronominit; harvinaisimpia ovat konjunktio-, kvanttori- ja adpositiokolligaatit. (Kuvion lyhenteet: adj = adjektiivit, adp = adpositiot, adv = adverbit, kva = kvanttorit, kon = konjunktiot, LA = lauseenalkuisuus, LL = lauseenloppuisuus, pro = pronominit, sub = substantiivit, ver = verbit.)



Kuvio 6.3. Kolligaatit tarkastelualueella V3–O2.

⁵⁶ *hyvin* vs. *kovin* $z = 2,218038$ ($p < 0,05$); *kovin* vs. *oikein* $z = 2,096117$ ($p < 0,05$).

⁵⁷ *hyvin* vs. *kovin* $z = 2,363603$ ($p < 0,05$); *hyvin* vs. *oikein* $z = 2,007781$ ($p < 0,05$).

6.2.2. Kolligationaalinen valinta positioittain tarkastelualueella

Positiokohtainen tarkastelu etenee vasemmista positioista oikeapuoleisiin positioihin. Kuvauksessa keskitytään astemääritteiden kolligationaalisten suhteiden eroihin; synonyymisyyttä korostavat tekijät ilmenevät yhtäläisenä kolligaattivariaationa, erottavat piirteet taas kolligaattien suhde-eroina. Yksittäisten kolligaattien osuudet on ilmoitettu positioittain liitteen 1 pylväsdiagrammeissa.

Position V3 yleisimmät kolligaatit ovat verbit, substantiivit ja lauseenalkuisuus. Näistä substantiivikolligaatti on selvästi yleisin *kovin*-määritteellä (29,0 %), lauseenalkuisuus taas *hyvin*- ja *oikein*-määritteillä (24,1 % ja 25,6 %). Astemääritteitä erottavat toisistaan tilastollisen testauksen perusteella kuitenkin ainoastaan lauseenalkuisuuden osuudet astemääriteparien *hyvin*–*kovin* ja *oikein*–*kovin* välillä:

Taulukko 6.12. V3-kolligaatit, jotka erottavat tilastollisin perustein astemääritteet toisistaan.

<i>hyvin</i> – <i>kovin</i>	<i>hyvin</i> – <i>oikein</i>	<i>kovin</i> – <i>oikein</i>
lauseenalkuisuus z = 2,18, p < 0,05	–	lauseenalkuisuus z = 2,06, p < 0,05

Selitys lauseenalkuisuuden muita pienemmälle osuudelle *kovin*-määritteellä tässä positiossa löytyy osittain kieltoverbillisten liittomuotojen sijoittumisesta vasemmanpuoleiseen tarkastelualueeseen. Osana verbien liittumotoja EI pidentää lauserakennetta hakusanasta vasemmalle päin, jolloin lauseenalkuisuus jää vähäisemmäksi tässä positiossa (esimerkkejä konkordanssin 6.8 riveillä 1–2). Toinen selittävä seikka on substantiivien muita suurempi osuus: vaikka niitä ei olekaan tilastollisesti merkitsevästi enemmän kuin muilla määritteillä tässä kohtaa kotekstia, on niiden osuus kuitenkin suuri ja vähentää osaltaan lauseenalkuisuutta. Substantiivit ovat *kovin*-määritteen kotekstissa usein subjektin funktiossa juuri tässä positiossa ennen kieltoverbien ja predikaattiverbien muodostamaa liittumotoa, jotka taas edeltävät *kovin*-määritettä (rivit 3–4). Lauseenalkuisuusesimerkkejä riveillä 5–9.

Konkordanssi 6.8. Esimerkkejä V3-kolligaaateista.

	V3	V2	V1	O1	O2		
1 in pora. Teillä	ei	taida	mennä	<i>kovin</i>	hyvin.	LL	Mä luulen
2 äiset avioparit	eivät	jaksaneet	enää	<i>kovin</i>	pitkään	istua	iltaa. Nuo
3 oreparvessa".	Aine	ei	ole	<i>kovin</i>	pitkä,	kuusi	lausetta
4 mille. Rajalla	jono	ei	näyttänyt	<i>kovin</i>	pitkältä	LL	mutta se e
5 puiston halki.	LA	Ne	ajoiivat	<i>hyvin</i>	kovaa,	LL	nainen pel
6 äneitä miehiä,	LA	mutta	vain	<i>hyvin</i>	harvat	heistä	olivat enä
7 dosta äänestä,	LA	joka	kuullaan	<i>hyvin</i>	harvoin	LL	vain silloi
8 opivia hetkiä.	LA	Kokouksissa	on	<i>oikein</i>	hyvä	torkahdella.	Harvoin n
9 n avioliittoon.	LA	Vanhan	kellarissa	<i>oikein</i>	hyvä	kaveri	pyysi kau

Positiossa V2 kolligaattien yleisyysjärjestys osoittautuu erilaiseksi, mutta nyt joukosta erottuu *oikein*: sen yleisin kolligaatti on sanaluokka substantiivit, sitten pronominit ja kolmanneksi lauseenalkuisuus; sen sijaan sekä *hyvin* että *kovin* saavat yleisimmiksi kolligaateikseen verbit, substantiivit ja pronominit (kuvio 2 liitteessä 1). Myös kolligaattien keskinäiset osuudet ovat astemääritteillä varsin erilaiset tässä asemassa (taulukko 6.13). *hyvin* poikkeaa ensinnäkin *kovin*-määritteestä kolmen kolligaatin saaman osuuden suhteen. Substantiivien ja pronomien osuudet (23,6 ja 18,2 %) ovat sen kolligaattien joukossa suuremmat kuin *kovin*-määritteellä (10,8 ja 11,4 %), ja vastaavasti verbien osuus (27,0 %) on taas huomattavasti pienempi kuin *kovin*-määritteellä (53,4 %). *oikein*-synonyymistaan *hyvin* eroaa verbien ja myös lauseenalkuisuuden osuuden suhteen: verbien osuus on *oikein*-määritteen V2-asemassa pienempi (14,0 vs. 27,0 %) ja lauseenalkuisuus taas suurempi (16,5 vs. 8,2 %) kuin *hyvin*-määritteellä. Viimein tarkasteltaessa jäljellä olevaa synonyymiparia eli *kovin*- ja *oikein*-määritteitä löytyy jälleen kolme kolligaattia, joiden osuuksien erot osoittautuvat tilastollisesti merkitsevästi erilaisiksi. Kuten kahta ensimmäistäkin paria, myös tätä paria erottaa verbikolligaatin osuus, ja vieläpä erittäin selvästi: *kovin*-määritteellä on verbejä tässä positiossa 53,4 prosenttia kaikista kolligaateista, ja *oikein*-määritteellä vastaava osuus on vain 14,0 prosenttia. Ero on siis hyvin huomattava, lähes 40 prosenttiyksikköä. Myös substantiivien ja lauseenalkuisuuden osuudet ovat selvästi erilaisia. Näitä on *oikein*-määritteellä tässä positiossa *kovin*-määritettä enemmän, sillä substantiivien osuus on 28,1 prosenttia ja lauseenalkuisuuden 16,5, kun osuudet ovat *kovin*-määritteellä 10,8 ja 4,5 prosenttia.

Taulukko 6.13. V2-kolligaatit, jotka erottavat tilastollisin perustein astemääritteet toisistaan.

<i>hyvin–kovin</i>	<i>hyvin–oikein</i>	<i>kovin–oikein</i>
verbit		verbit
$z = 5,97, p < 0,001$	verbit	$z = 6,89, p < 0,001$
substantiivit	$z = 2,89, p < 0,01$	substantiivit
$z = 3,51, p < 0,001$	lauseenalkuisuus	$z = 3,83, p < 0,001$
pronominit	$z = 2,58, p < 0,01$	lauseenalkuisuus
$z = 2,02, p < 0,05$		$z = 3,47, p < 0,001$

Kolligaateista verbien sanaluokka erottelee kaikki astemääritteet toisistaan. Verbit ovat ylivoimaisesti yleisin kolligaatti *kovin*-astemääritteellä tässä positiossa, ja myös selvästi synonyymeja erottava tekijä. Selitys määritteen V2-verbikolligaatin suureen osuuteen on selvä: sen aiheuttaa kieltoverbin EI esiintyminen hyvin usein tässä positiossa. *kovin* saa kieltoverbin tähän asemaan 49 kertaa, mikä on 52,1 prosenttia kaikista sen verbeistä tässä positiossa (konkordanssissa 6.11 rivit 1–4), muilla määritteillä kieltoverbin esiintymät ovat olemattomat. Kieltoa ilmaisevan apuverbin sijoittuminen tähän positioon selittää siis *kovin*-määritteen verbikolligaatin osuuksien eron kahteen muuhun astemääritteeseen, mutta mikä selittää *hyvin*- ja *oikein* määritteen verbikolligaattien osuuksien eron, joka sekin osoittautui tilastollisesti merkitseväksi? *hyvin*-määritteelle on tyypillistä, että sitä edeltävässä koteksissa oleva verbikolligaatti sijoittuu heti vieruskolligaatiksi (positio V1, jossa osuus on 42,9 prosenttia kaikista kolligaateista), mutta se voi sijoittua myös V2-kolligaa-

tiksi, jolloin sen ja astemääritteen väliin tulee jokin lauseenjäsen, kuten objekti (konkordanssissa 6.9 rivi 5), adverbiaali (rivi 6) tai predikaattiverbin liittomuodon pääverbi (rivi 7). *oikein*-määritteen sisältävässä lauseessa sen sijaan verbikolligaatti on selvästi useammin vieruskolligaattina (50,4 %) kuin V2-kolligaattina (14,0 %). Sen vasemmanpuoleinen koteksti ei siten suosi rakennetta, jossa verbikolligaatin ja astemääritteen väliin tulisi jokin muu kolligaatti, eikä myöskään rakenteita, joissa sitä edeltäisi verbiketju (vain kaksi tapausta, rivit 8 ja 9), ja siksiäpä verbi ei sijoitu useinkaan V2-asemaan. Substantiivikolligaatin (*hyvin-kovin* ja *oikein-kovin*) ja pronominikolligaatin (*hyvin-kovin*) osuuksien ero on ainakin osittain selitettävissä verbikolligaatin osuuksilla. Kun *kovin*-määritteen V2-positiossa on runsaasti kieltoverbejä, ovat lauseessa usein subjektina toimivat substantiivit ja pronominit kauempana astemääritteestä kun muilla astemääritteillä, useimmiten positiossa V3, mikä tuli jo ilmi edellä. Tämä vähentää siten substantiivi- ja pronominikolligaattien osuutta V2-positiossa. Muilla määritteillä etenkin substantiivikolligaatin osuus on tässä asemassa lähes yhtä suuri (*hyvin*) tai suurempi (*oikein*) kuin verbikolligaatin osuus. Myös *oikein*- ja *kovin*-määritteiden välillä pronominikolligaatin osuuden ero on melko suuri, mutta tässä aineistossa se ei osoittautunut tilastollisesti merkitseväksi.

Konkordanssi 6.9. Esimerkkejä V2-kolligaaeteista.

	V3	V2	V1		O1	O2	
1	entekevää.	Se	ei	ole	<i>kovin</i>	kaukana	sinun nykyisestä
2	kasti, mutta	minä	en	ollut	<i>kovin</i>	innostunut.	LL Muut ihmet
3	, että lasten	vuoksi	emme	voisi	<i>kovin</i>	myöhään	viipyä, mihin myö
4	että tämä	hallitus	ei	taida	<i>kovin</i>	kauan	enää paikallaan
5	tää. Ja siinä	puhuessani	muistin	kissan	<i>hyvin</i>	tarkoin.	LL Minä olin t
6	mies sanoi.	Sinulla	oli	silloin	<i>hyvin</i>	pitkä	tukka, jolle satoi
7	sä takaisin.	LA	On	tullut	<i>hyvin</i>	hiljaista,	LL linnut minu
8	ymmärtää.	LA	Alan	ymmärtää	<i>oikein</i>	hyvin.	LL Ne eivät lu
9	oiksa tulla?	Kullervo	ei	voi	<i>oikein</i>	hyvin.	LL Meillä jatk
10	enää kauan	kestä,	LA	ottakaa	<i>oikein</i>	hyvä	asento ja rentoutu
11	eneeseen ja	sanoivat:	LA	Jos	<i>oikein</i>	paha	paikka tulee, Cale
12	tää ja meitä	paleli.	LA	Oli	<i>hyvin</i>	hiljaista.	LL En muista
13	estä järven	pinnasta;	LA	oli	<i>hyvin</i>	lämmin.	LL Martikaine
14	mistä on	kysymys.	LA	Olin	<i>hyvin</i>	yllättynyt	Sarin kehrittelem

Kuten edellisessäkin positiossa, myös asemassa V2 lauseenalkuisuus erottaa *oikein*-määritteen synonyymeistaan. *kovin*-määritteellä lauseenalkuisuuden osuus on minimaalinen, vain 4,5 prosenttia, ja *hyvin*-määritteelläkin vain 8,2 prosenttia. Lauseenalkuisuuskolligaatti näyttää siten olevan tyypillisempi vasemmanpuoleinen kolligaatti juuri *oikein*-määritteelle (16,5 %). V2-lauseenalkuisuuskolligaatit selittyvät ainakin osin kehotuksista tai muista repliikeistä, joissa on lauseessa ensimmäisenä (siis V1-positiossa) verbikolligaatti (rivit 10, 11). *hyvin*-määritteen kotekstissa

oleva lauseenalkuisuuskolligaatti on taas erilaisissa rakenteissa, pääosin lyhyissä tilalauseissa (rivit 12, 13) tai persoonapäätteisen *olla*-verbin aloittamissa lyhyissä lauseissa, joissa on astemääritteen määrittämä predikatiivi (rivi 14). Nämä rakenteet ovat kuitenkin niin harvinaisia, että lauseenalkuisuuskolligaatti ei ole tavallinen.

Määritteiden vasemmista vieruskolligaaateista (asema V1) on ylivoimaisesti suurin ryhmä kaikilla astemääritteillä verbit. Verbien osuus tässä positiossa on selitettävissä helposti: ne edeltävät yleensä joko astemääritteen sisältävää adverbiaalilauseketta (kuten konkordanssissa 6.10 rivillä 1), tai toimivat kopulana astemääritteen sisältävän predikatiivilausekkeen edellä (rivi 2). Seuraavaksi yleisimmät sanaluokat ovat tässä positiossa adverbit ja substantiivit; konjunktioiden, pronomien ja muiden sanaluokkien osuudet jäävät vähäisiksi, samoin kuin lauseenalkuisuus.

Vaikka kolligaattien järjestys ei tässäkään positiossa poikkea paljon astemääritteiden välillä, on kolligaattien osuuksissa jälleen eroja. Ensinnäkin *hyvin* poikkeaa *kovin*-määritteestä verbien ja konjunktioiden määrän osalta. Sillä on tässä positiossa verbejä vähemmän ja konjunktioita enemmän kuin *kovin*-määritteellä: verbien osuus on *hyvin*-astemääritteellä 42,9 prosenttia ja konjunktioiden osuus 8,2 prosenttia, kun taas vastaavat prosenttimäärät ovat *kovin*-määritteellä noin 58,5 ja 2,3. Erot ovat tilastollisesti merkitseviä. *oikein*-määritteestä *hyvin* poikkeaa taas lauseenalkuisuuden suhteen, sillä niitä on *hyvin*-määritteellä vain 6,3 prosenttia tässä asemassa, kun *oikein*-määritteellä niitä on yli kaksin verroin, 14,0 prosenttia. Lauseenalkuisuus on myös *kovin*- ja *oikein*-määritteitä erottava seikka: edellisellä on lauseenalkuisuutta vain 4,5 prosenttia sen kaikista kolligaaateista. Muut kolligaatit eivät erota synonyymeja toisistaan. (Taulukko 6.14.)

Taulukko 6.14. V1-kolligaatit, jotka erottavat tilastollisin perustein astemääritteet toisistaan.

<i>hyvin–kovin</i>	<i>hyvin–oikein</i>	<i>kovin–oikein</i>
verbit		
z = 3,39, p < 0,001	lauseenalkuisuus	lauseenalkuisuus
konjunktiot	z = 2,69, p < 0,01	z = 2,90, p < 0,01
z = 2,67, p < 0,01		

Kvantitatiivisen eron lisäksi lauseenalkuisuuteen liittyy myös kvalitatiivisia eroja, jotka voivat selittää myös määrälliset erot. Etenkin *kovin*- ja myös *hyvin*-määritteellä lauseenalkuisuutta esiintyy silloin, kun lauseen objekti (konkordanssi 6.10 rivi 3) tai yleisemmin adverbiaali (rivit 4 ja 5) on fokusoitu lauseen alkuun. *hyvin*-määritteen yhteydessä on myös joitakin elliptisiä rakenteita, joiden tulkinta täydentyy kontekstista ja joissa määrite on rakenteen alussa (6). Myös *oikein* voi olla fokusoidun aineksen määritteenä (7) ja elliptisissä rakenteissa (8), mutta lauseenalkuisuus sijoittuu määritettä ennen erityisesti silloin, kun määrite esiintyy osana henkilöiden vuoropuhelua. Tällaisissa tapauksissa *oikein* voi olla osa kehotusta tai toivomusta, jotka ovat lauseina vaillinaisia (9). Nämä tapaukset lisäävät osaltaan *oikein*-määritteen lauseenalkuisuuskolligaattien määrää.

Konkordanssi 6.10. Esimerkkejä V1-kolligaateista.

	V3	V2	V1		O1	O2	
1	tai	leipä	loppuu	<i>hyvin</i>	äkkiä,	LL	Petteri oli väittäit
2	vierailu	ei	olisi	<i>kovin</i>	hyvä	juttu.	Älä kuule, Huj
3	ja	söi.	LA	<i>Kovin</i>	merkillistä	ateriaa	Hartikainen ei pöytään ollut saanut loihdituksi.
4	kohde	odottaisi.	LA	<i>Kovin</i>	röyhkeäksi	hän	ei varmaankaan uskaltaisi heittäytyä.
5	oudoksuen	taivaalle.	LA	<i>Hyvin</i>	korkealla	kiiti	pilvenhaituvia huimaa vauhtia, vaikkei maan
6	marokko-laista	punaviiniä.	LA	<i>Hyvin</i>	punaista.	LL	Hiukan lämmintä. Sen maku toi mieleen muita
7	hautomisen	aika.	LA	<i>Oikein</i>	tarkasti	ko	katto, näki linnun pyöreät mustat silmät kiilumassa
8	sormen	anuksessaan.	LA	<i>Oikein</i>	hyvin	tunsi.	Eipä ollut aikaisemmin tuntenutkaan
9	viiltävä	ironia.	LA	<i>Oikein</i>	hyvää	matkaa,	minä sanoin. - Pitäkää h

hyvin-määritteen konjunktio-kolligaatit ovat yhtä poikkeusta lukuun ottamatta rinnastuskonjunktioita. Määritteelle on tyypillistä, että se esiintyy rakenteissa, joissa on rinnastettu kaksi tai useampi adjektiivi (konkordanssi 6.11 rivi 1) tai adverbi (2), ja joista ainakin jälkimmäistä määrittää *hyvin*. Toisinaan rinnastuskonjunktio aloittaa myös lauseen (3). Molemmilla muilla astemääritteillä tällaiset rakenteet ovat harvinaisia, sen sijaan niille on tyypillistä, että määrite sijaitsee ennen rinnastusta (4 ja 5). *hyvin*-määritteen verbikolligaatin osuutta pienentävät muiden kolligaattien (paitsi pronomien) hieman suuremmat osuudet suhteessa *kovin*-määritteeseen.

Konkordanssi 6.11. Esimerkkejä V1-konjunktio-kolligaateista.

	V3	V2	V1		O1	O2	
1	tan kaasona on tumma, kaunis	ja	pelokas	ja	<i>hyvin</i>	surullinen	tyttö. Hän oli
2	Saatana, Marsipaani sanoi	hyvin	hitaasti	ja	<i>hyvin</i>	hiljaa.	LL Anna antee
3	Veijo Salmen	kanssa.	LA	Mutta	<i>hyvin</i>	pian	Markkainen sai huomata, että
4	oli kiva	seurata	LA	miten	<i>oikein</i>	vanha	ja kömpelö mies yrittää liimata
5	niin, että	meidän	murheemme	ovat	<i>kovin</i>	pieniä	ja ohimeneviä.

Vaikka näiden muiden kolligaattien osuudet eivät konjunktio-kolligaattia lukuun ottamatta osoittaudu tilastollisesti merkitsevästi erilaisiksi, ne osaltaan kuitenkin pienentävät verbikolligaatin osuutta.

Edellä luvussa 5.3.2 käsitellystä astemääritekategoriaa koskevasta katsauksesta on jo pääteltävissä, millaisia sanaluokkia esiintyy astemääritteiden oikeanpuoleisessa ympäristössä. Koska kyse on etumääritteistä, kolligaatiopositiossa O1 eli oikeanpuoleisina vieruskolligaatteina sijaitsevat astemääritteiden pääsanat. Täten voidaan olettaa jo ennen korpuksen analyysia, että astemääritteiden O1-kolligaatteina esiintyy ainakin adjektiiveja, adverbeja ja kvanttoreita, mahdollisesti myös adpositioita. Kuitenkaan eri astemääritteiden tai edes eri astemääriteryhmien osalta ei ole tehty tutkimusta, joka paljastaisi näiden pääsanojen sanaluokkien keskinäiset osuudet. Seuraava katsaus valaisee tätä asiaa selvittämällä tutkittavana olevien kolmen astemääritteen kolligaatit eli pääsanat positiossa O1.

Adjektiivikolligaatin osuus on selvästi suurin kaikilla määritteillä tässä positiossa; osuus vaihtelee *hyvin*-määritteen 66,5 prosentista *oikein*-määritteen 55,4 prosenttiin (kuvio 4 liitteessä 1). Adverbit ovat seuraavaksi yleisin kolligaattiryhmä ja sen jälkeen tulevat kvanttorit. Adpositiorakenteiden osuus on häviävän pieni: niiden osuus jää alle kahden prosentin kaikilla astemääritteillä. Korpuksen perusteella näyttää siltä, että kolligaattien keskinäinen yleisyysjärjestys ei erota synonyymisia astemääritteitä toisistaan, sillä järjestys on kaikilla määritteillä sama. Määritteiden saamien kolligaattien osuudet kuitenkin vaihtelevat. Erityisesti syntagmat <*oikein* + adverbikolligaatti> ja <*kovin* + kvanttorikolligaatti> näyttävät erottuvan frekvensteiksi. Seuraava tarkastelu keskittyykin syntagmakombinaatioihin jälleen astemääritepareittain.

Ensinnäkin astemääritteiden *hyvin* ja *kovin* kolligaattien osuudet ovat melko samanlaiset muiden kolligaattien paitsi kvanttoreiden osalta: *hyvin*-määritteellä on kvanttorikolligaatin osuus vain 4,8 prosenttia, mutta *kovin*-määritteellä osuus on 15,3 prosenttia (konkordanssissa 6.12 rivit 1–3). Muiden kolligaattien osuudet eivät erota näitä synonyymeja toisistaan. (Taulukko 6.15.) Kun tarkastelussa siirrytään vertaamaan *hyvin*-astemääritettä *oikein*-määritteeseen, lisääntyy synonyymeja erottavien tekijöiden määrä yhdellä. Astemääritteitä erottaa toisistaan ensiksikin adjektiivien osuus: adjektiivien osuus on *oikein*-astemääritteellä 55,4 %, kun *hyvin*-määritteellä se on 66,5 %. Myös adverbien osuuksien ero (*hyvin*-määritteellä 27,8 % ja *oikein*-määritteellä 39,7 %) osoittautuu näiden astemääritteiden välillä selvästi erilaiseksi. Kvanttoreiden ja adpositioilmausten osuudet ovat näiden kahden astemääritteen sanaluokkakolligaateista likimain samoja, eivätkä siten osoittaudu erottaviksi tekijöiksi. Konkordanssissa 6.12 riveillä 4–6 on esimerkkejä adjektiivi-, adverb- ja adpositiokolligaateista. Viimeisen parin eli *kovin*- ja *oikein*-astemääritteiden vertailusta selviää, että adverbien osuus on jälleen synonyymeja erottava tekijä. Osuudet ovat 22,7 prosenttia (*kovin*) ja 39,7 prosenttia (*oikein*). Lisäksi myös kvanttorikolligaattien osuuksien ero (*kovin*-määritteellä 15,3 prosenttia ja *oikein*-määritteellä 5,0 prosenttia) erottaa synonyymit toisistaan. Muiden sanaluokkakolligaattien osuuksissa ei ole merkitseviä eroja.

Tilastolliseen merkitsevyyteen perustuvan analyysin mukaan astemääriteparin *hyvin*–*kovin* jäsenet näyttäisivät olevan keskenään synonyymisempia kuin astemääriteparit *hyvin*–*oikein* ja *kovin*–*oikein*, sillä ensiksi mainitun parin jäseniä erottaa toisistaan vain yhden kolligaatin osuuksissa ilmenevä ero, kun taas jälkimmäisiä erottavat kahden kolligaatin osuuksissa olevat erot.

Taulukko 6.15. O1-kolligaatit, jotka erottavat tilastollisin perustein astemääritteet toisistaan.

<i>hyvin–kovin</i>	<i>hyvin–oikein</i>	<i>kovin–oikein</i>
	adjektiivit	adverbit
kvanttorit z = 4,12, p < 0,001	z = 2,19, p < 0,01	z = 3,14, p < 0,01
	adverbit	kvanttorit
	z = 2,43, p < 0,01	z = 2,80, p < 0,01

Konkordanssi 6.12. Esimerkkejä oikeanpuoleisista vieruskolligaateista.

	V3	V2	V1		O1	O2	
1	kiitteli	työtään;	LA	<i>kovin</i>	monta	naulaa	se ei kirveen hamara
2	kiinnittänyt	silloin	poikaan	<i>kovin</i>	paljon	huomiota	eikä ehkä poika min
3	jonka	osasin.	LA	<i>hyvin</i>	vähän	se	oli, mutta taisi pitää
4	ollut	kaikesta	huolimatta	<i>hyvin</i>	viehättävä	vaikka	vähän kalpea. Oli oj
5	näki	kaiken	sen	<i>oikein</i>	selvästi	mutta	ei ymmärtänyt.
6	LA	toi	kasvonsa	<i>hyvin</i>	lähelle	minun	kasvojani ja kysyi

Seuraavaksi tarkastellaan viimeisen jäljellä olevan position eli O2-aseman kolligaattikirjoa. Myös tästä asemasta löytyy muuttujia, jotka sekä yhdistävät astemääritteitä toisiinsa että toisaalta erottavat ne toisistaan. Ensinnäkin lauseenloppuisuus on tässä asemassa yleisin kolligaatti: sen osuus vaihtelee *kovin*-määritteen 40,9 prosentista *hyvin*-määritteen 52,3 prosenttiin (konkordanssissa 6.13 rivit 1 ja 2). Seuraavaksi yleisin kolligaatti on kaikilla sanaluokka substantiivit, kuten rivillä 3, jossa astemäärite määrittää substantiivilausekkeessa olevaa adjektiivia. Kolmanneksi tavallisin kolligaatti sen sijaan hieman vaihtelee astemääritteittäin: *oikein*- ja erityisesti *kovin*-määritteellä se on verbi (rivi 4), mutta *hyvin*-määritteellä mikään jäljellä olevien sanaluokkien osuus ei ole yhtä selvästi muita suurempi.

Konkordanssi 6.13. Esimerkkejä O2-kolligaateista.

	V3	V2	V1		O1	O2	
1	myös	Harlekiini	on	<i>hyvin</i>	mukava.	LL	Hän riemastuttaa sy
2	LA	Kiitos	nyt	<i>oikein</i>	paljon,	LL	Sarlotta sanoo.
3	auttoi.	LA	Kaksi	<i>kovin</i>	erilaista	naista	olivat ystäväystyneet.
4	nainen	ei	voi	<i>kovin</i>	luontevasti	liittää	autoa itseensä genit

Vaikka lauseenloppuisuus on tässä asemassa yleisin kolligaatti kaikilla määritteillä, osoittautuu sen suhteellinen osuus kuitenkin astemääritteitä erottavaksi tekijäksi: *hyvin*-astemääritteellä sen osuus (52,3 prosenttia) on nimittäin *kovin*-astemääritteen osuutta (40,9 prosenttia) tilastollisesti merkittävästi suurempi. *oikein* ja *hyvin* ja myöskään *oikein* ja *kovin* eivät eroa tämän muuttujan suhteen

toisistaan.⁵⁸ Toinen erottava tekijä on verbien osuus. *hyvin*-astemäärityksen verbikolligaatin osuus on tilastollisesti merkitsevästi suurempi kuin vastaavat osuudet *kovin*- ja *oikein*-määrittelyillä, jotka taasen eivät eroa toisistaan⁵⁹. Muiden kolligaattien osuuksissa ei ole tilastollisesti merkitsevää eroa, joten kolligaattien osuuksien suhteen synonyymit eivät kovin paljon eroa toisistaan tässä positiossa. Eniten näyttävät eroavan *hyvin* ja *kovin*, joita erottavat sekä verbien määrä että lauseenloppuisuuden määrä. *hyvin* ja *oikein* eroavat vain verbien osuuden suhteen, ja *kovin* ja *oikein* eivät tämän muuttujan perusteella eroa ollenkaan. Tilanne on siis tässä positiossa ainakin osittain päinvastainen kuin asemassa O1: vieruskolligaatti-analyysin mukaan synonyymisimpia olivat juuri *hyvin* ja *kovin*, *kovin*–*oikein*-paria taas erottivat kahden sanaluokkakolligaatin ja lisäksi tapaa ja aikaa ilmoittavien adverbikolligaattien suhteelliset osuudet. Näyttää siis siltä, että määrittelyiden synonyymisuus voi vaihdella positioittain. Kvantitatiivisten kriteereiden perusteella tehty analyysi on esitetty tiivistetysti taulukossa 6.16:

Taulukko 6.16. O2-kolligaatit, jotka erottavat tilastollisin perustein astemääritykset toisistaan.

<i>hyvin</i> – <i>kovin</i>	<i>hyvin</i> – <i>oikein</i>	<i>kovin</i> – <i>oikein</i>
verbit	verbit	–
z = 4,94, p < 0,001	z = 2,94, p < 0,01	–
lauseenloppuisuus		
z = 2,46, p < 0,01		

Tässä yhteydessä on jälleen tähdellistä tarkastella, millaisia syitä on mahdollista löytää verbien kvantitatiivisten erojen takaa. Selitys verbikolligaatin osuuksien eroihin löytyy astemääritysten ja niiden pääsanojen kytkeytymisestä verbirakenteisiin. Ensiksikin *kovin* ja sen pääsana esiintyvät aineistossa usein verbiryhmän muodostavien verbien välissä (konkordanssissa 6.14 rivit 1–3). Näissä rakenteissa verbiryhmän jälkimmäinen verbi on sijoittunut juuri O2-asemaan. Verbiryhmiä ovat sekä liittomuodot, jotka muodostuvat apuverbeistä (*olla*-verbi ja kieltoverbi *ei*) ja pääverbistä, että verbiketjut, joiden ensimmäisenä jäsenenä voi olla muun muassa modaalin verbi (*voida*, *sallia*) tai vaikkapa aistihavaintoa kuvaava verbi (*kuulua*, *näkyä*) (ks. Hakulinen ja Karlsson 1979: 230-35). Vilkunan (1996: 253, 277) mukaan sekä liittomuodoille että verbiketjuille on ominaista, että niiden osat eivät pysy lauseessa vierekkäin, jolloin niiden osien välissä voi sijaita lauseen muita jäseniä. Usein ne ovat adverbiaaleja (*Uolevi olisi sunnuntaina mielellään lähtenyt mukaan*, liittomuoto), (*Minä lakkasin ylipäätään viikonloppuisin käymästä ulkona*, verbiketju), mutta välissä voi sijaita niin subjekti (*Sen nyt olisi lapsikin ymmärtänyt*) kuin objektikin (*Uolevi haluaa sen tehdä*). Tavallista tämä on silloin, kun apuverbi tai verbiketjun finiittiverbi aloittaa lauseen: **Ei Uolevi sitä tahallaan tehnyt.** (Vilkuna mp., esimerkit Vilkinson.)

⁵⁸ *hyvin* vs. *kovin* z = 2,462960 (p < 0,05); *hyvin* vs. *oikein* z = 1,921423 (p > 0,05); *kovin* vs. *oikein* z = 0,213111 (p > 0,05).

⁵⁹ *hyvin* vs. *kovin* z = 4,944405 (p < 0,001); *hyvin* vs. *oikein* z = 2,935496 (p < 0,01); *kovin* vs. *oikein* z = 1,267973 (p > 0,05).

Juuri adverbiaalien kiilautuminen verbiryhmien osien väliin on tyypillistä niissä tapauksissa, joissa *kovin* saa verbikolligaatin O2-positioon: peräti 83,3 prosentissa kaikista tapauksista, joissa verbi on tässä positiossa, on astemäärityksen sisältävä adverbiaalilauseke verbiryhmän jäsenten välissä. Tapaukset ovat sekä liittomuotoja (konkordanssin 6.14 rivit 1–2) että verbiketjuja (rivi 3), ja useimmiten adverbiaalilausekkeen pääsanana on OSMA eli objektinsijainen määrän adverbiaali (rivi 1), mutta myös aikaa (rivi 2) ja tapaa (rivi 3) ilmoittava adverbiaali (OSMA:ista mm. Hakulinen–Karlsson 1979: 216–218; Tuomikoski 1978). Adverbiaalien kiilautuminen verbien väliin voi osittain selittää myös *oikein*-määrityksen verbikolligaattien suhteellisen taajan esiintymisen, sillä nämä muodostavat verbikolligaattitapauksista kolmanneksen. Tätä yleisempi on kuitenkin rakenne, jossa tekemisen tapaa ilmoittava adverbiaali ja sen saama astemääritykset ovat fokusoidussa asemassa ennen verbiä (rivi 4):

Konkordanssi 6.14. Verbiryhmiä (1–3) ja adverbiaalin fokuointia (4).

	V3	V2	V1		O1	O2	
1	joista	hänelle	oli	<i>kovin</i>	vähän	ollut	hyötyä. Elinaa ei s
2	osasto	ei	ole	<i>kovinkaan</i>	pitkään	kuulunut	luonnonvalinnan t
3	nainen	ei	voi	<i>kovin</i>	luontevasti	liittää	autoa itseensä geni
4	jos	isossa	kaupungissa	<i>oikein</i>	sitkeästi	pidetään	huhua hengissä, se

oikein-määrityksen O2-positiossa olevien verbien määrä on kuitenkin niin vähäinen (14 esiintymää), että sen perusteella on mahdotonta sanoa selvää syytä siihen, miksi verbikolligaatti osoittautuu aineistossa *hyvin*-määrityksen verbikolligaattia yleisemmäksi. *hyvin*-astemäärityksellä edellisen kaltaisista rakenteista kumpikaan ei ole erityisen tavallinen, eikä sellaisista kombinaatioista, joissa verbi on O2-asemassa, löydy mitään pelkästään *hyvin*-määritykselle tyypillistä esiintymisympäristöä. Lauseenloppuisen aseman muita suurempi osuus *hyvin*-määrityksen O2-kolligaattien joukossa selittyy pääasiassa verbikolligaatin osuuden pienuudella: *hyvin*-määrityksen verbikolligaattien osuus on muiden astemääritysten vastaavaa osuutta pienempi, ja toisaalta lauseenloppuisuus tässä asemassa on muita selvästi suurempi.

6.2.3. Kolligaattien klusteroituminen

Kollokaattianalyysissä luvussa 6.1.3 huomattiin, että astemääritykset esiintyvät omanlaisissaan sanaklustereissa. Kolligaattien klusteroituminen on kollokaattien klusteroitumista odotuksenmukaisempaa, sillä lauseiden kieliopilliset ominaisuudet, kuten lauseke- ja sanajärjestysmahdollisuudet, rajoittavat kolligaattien esiintymisjärjestystä. Siten on myös oletettavaa, että kolligaattiklustereiden esiintymätaajuuDET ovat suuremmat kuin sanaklustereiden. Seuraavassa tarkastellaan kolligaattiklustereita astemäärityksittäin siten, että jokaisen astemäärityksen konkordanssista on etsitty kolmen kolligaatin mittaiset klusterit, joista tarkastellaan kolmea yleisintä

klusterirakennetta. Näiden avulla pyritään luomaan yleistyksiä yleisimmistä kombinaatioista ja eräänlaisia perussyntagmaja, joiden osana määritteet esiintyvät tyypillisimmin. *hyvin*-määritteen kolmiosaiset kolligaattiklusterit ovat taulukossa 6.17 esitetyn kaltaiset.

Taulukko 6.17. *hyvin*-määritteen kolligaattiklusterit.

	V2	V1	noodi	O1	O2	f
1	PRO	VER	<i>hyvin</i>			53
2	SUB	VER	<i>hyvin</i>			49
3	VER	ADV	<i>hyvin</i>			24
4		VER	<i>hyvin</i>	ADJ		113
5		ADV	<i>hyvin</i>	ADJ		36
6		VER	<i>hyvin</i>	ADV		35
7			<i>hyvin</i>	ADJ	LL	117
8			<i>hyvin</i>	ADJ	SUB	75
9			<i>hyvin</i>	ADV	LL	60

Kolligaattien klusteroitumiskaavio kuvastaa selkeää rakennetta, jossa *hyvin*-määritettä käytetään: klustereissa positiossa V2 on tavallisesti joko pronomini tai substantiivi, seuraavassa positiossa verbi, oikeanpuoleisessa kotekstissa positiossa O1 tyypillisimmin adjektiivi ja viimeisessä asemassa lauseenloppuisuus. Kolligaateista muodostuu siten hypoteettinen viisiosainen klusteri

<PRO/SUB VER *hyvin* ADJ LL>,

joita myös tavataan aineistosta: pronomiinalkuinen klusteri esiintyy 23 kertaa ja on yleisin viisijäsenisistä klustereista, substantiivialkuinen löytyy lähes yhtä usein, 21 kertaa. Näiden yhteenlaskettu osuus on 12,5 prosenttia koko aineistosta, eli yli joka kymmenes *hyvin*-määritteen esiintymistä on osa jompaakumpaa klusteria. Näille klustereille on myös tyypillistä, että asemassa V3 on usein lauseenalkuisuus. Tyypillisimmäksi koko tarkastelualueen kattavaksi kolligaattiklusteriksi on abstrahoitavissa rakenne

<LA PRO VER *hyvin* ADJ LL>,

joka löytyy aineistosta yhteensä 15 kertaa (4,3 %). Nämä ovat nähtävissä konkordanssista 6.15 (s. 118).

Konkordanssi 6.15. *hyvin*-määritteen yleisin kolligaattisyntagma <LA PRO VER *hyvin* ADJ LL>.

1	Oletko sinä tehnyt aviorikosta? <i>Hän näytti hyvin uteliaalta</i> . -- En, minä sanoin.
2	ulkopuolelle nopein pitkin hypyin. <i>Hän oli hyvin laiha</i> , vihertävä puku oli repaleinen
3	luonaan pizzerian yläpuolella. <i>Hän oli hyvin uskonnollinen</i> , kuten hänen iässään
4	- Miten te tunnette hänet. - <i>Minä olen hyvin väsynyt</i> . - Me lopetamme ihan kohta
5	tutkimme poikanne murhaa. - <i>Minä olen hyvin väsynyt</i> . Kysykää siltä huoralta.
6	Oli. - Te olette varma siitä. - <i>Minä olen hyvin väsynyt</i> . Olen varma. Rouva huokasi
7	varmasti? Se on vuorenvarmaa. <i>Minä olen hyvin voimakas</i> . Minulle ei pärjää kukaan!
8	pyysivät Maisaa lopettamaan. <i>Minusta oli hyvin epätodennäköistä</i> , että sihteerini
9	kiinni? Hän sanoi. Tällä helteellä? <i>Ne ovat hyvin takertuvia</i> , minä sanoin. Vaikeita
10	, koska minä olen talouspäällikkö. <i>Se oli hyvin loukkaavaa</i> . Antero kertoi kuulleen
11	pyhimystä, Roccosta tulee gangsteri. <i>Se oli hyvin todennäköistä</i> , mikäli geeniperintö
12	juuri olisi satanut, eikö totta? Niin. <i>Se on hyvin kaunis</i> , Gamma sanoi innottomasti
13	olkapäillä mitä tehdään. - Tuota... <i>Tää on hyvin ystävällistä...</i> - Ei mitään. Mie oon
14	uuteen tietooni. Sanoin itselleni: <i>Tämä on hyvin opettavaista</i> , vaikka en mitenkään
15	kuin varsinaisesti seksuaalinen. <i>Tämä on hyvin yleistä</i> , mutta naaraspuolinen

Kolligaattiklustereiden analyysi tarkentaa kuvaa niistä tyypillisistä syntagmaista, joissa *hyvin*-määritettä käytetään. *hyvin*-määritteen *kollokaattiklustereiden* käsitteilyn yhteydessä nousi taajimmin esiintyväksi klusteriksi rakenne <SE OLLA *hyvin*>, ja yleisiä olivat myös muut klusterit, joissa positiossa V2 oli pronomini. Taajaan esiintyväksi erottui myös klusterirakenne <OLLA *hyvin* adjektiivikollokaatti>. Nämä molemmat rakenteet toistuvat kolligaattiklustereina, kuten edellä nähtiin, ja ovat taajimmin esiintyviä rakenteita. Kolligaattiklustereiden analyysi tuo kuitenkin esiin myös sellaista tietoa, mitä kollokaattien analyysi ei tavoittanut. Taulukossa 6.8 (s.102), johon on lueteltu taajimmin esiintyvät kollokaattiklusterit, ei esiinny ainutkukaan substantiivikollokaattia positiossa V2, mikä voisi luoda oletuksen, että substantiiveja ei ole mukana klusterirakenteissa. Näin tietysti onkin kollokaatteja tarkasteltaessa, mutta ei suinkaan astemääritteen ja kolligaattien muodostamien syntagmojen tasolla, onhan klusteri <SUB VER *hyvin* ADJ LL> lähes yhtä yleinen kuin pronominillakin alkava rakenne. Substantiivien puuttuminen kollokaattiklustereista selittyy sillä, että substantiiveista ei löydy koviin taajaan esiintyviä kollokaatteja, kuten selvisi luvussa 6.1.1.2. Lisäksi kolligaattien klusteroitumista tarkastelemalla tavoitetaan myös tieto lauseaseman osuudesta rakenteissa, mikä osoittautuu myös merkittäväksi.

Seuraavaksi tarkasteltavan *kovin*-määritteen kolmiosaiset klusterirakenteet on lueteltu taulukkoon 6.18. Samoin kuin edellä myös *kovin*-määritteen kolligaattiklustereista hahmottuu selvästi yksi rakenne, jossa määritettä käytetään. Rakenne on *hyvin*-määritteen rakenteen kanssa samankaltainen lukuun ottamatta V2-positiossa olevaa kolligaattia, joka on tyypillisimmin verbi. Oletettavaksi peruskluusteriksi muodostuu siten rakenne

<VER VER *kovin* ADJ LL>,

joka on aineistonkin perusteella yleisin viisijäseninen klusteri. Se esiintyy korpuk- sessa yhteensä 18 kertaa, mikä vastaa 10,2 prosenttia kaikista *kovin*-määritteen

esiintymistä. Näin ollen joka kymmenettä *kovin*-määritettä on käytetty nimenomaan tässä kolligationaalisessa rakenteessa. Jatkettaessa rakennetta aina asemaan V3 asti selviää, että *hyvin*-määritteestä poiketen tästä asemasta löytyy useimmin substantiivikolligaatti (n = 9; 5,1 %) eikä lauseenalkuisuus (konkordanssi 6.16):

<SUB VER VER *kovin* ADJ LL>.

kovin-määritteen klusterirakenteissa kieltoverbi EI sijoittuu tyypillisesti V2-positioon, joka erottaa klusterirakenteita tarkasteltaessa *hyvin*- ja *kovin*-määritteet toisistaan.

Taulukko 6.18. *kovin*-määritteen kolligaattiklusterit.

	V2	V1	noodi	O1	O2	f
1	VER	VER	<i>kovin</i>			51
2	VER	ADV	<i>kovin</i>			19
3	PRO	VER	<i>kovin</i>			14
4		VER	<i>kovin</i>	ADJ		63
5		VER	<i>kovin</i>	ADV		20
6		ADV	<i>kovin</i>	ADJ		18
7			<i>kovin</i>	ADJ	LL	55
8			<i>kovin</i>	ADJ	SUB	28
9			<i>kovin</i>	ADV	LL	14

Konkordanssi 6.16. *kovin*-määritteen yleisin kolligaattisyntagma <SUB VER VER *kovin* ADJ LL>.

1	eikkailuni kuoreparvessa.” <i>Aine ei ole kovin pitkä</i> , kuusi lausetta vain. Mutta to
2	ötäni. Ensiksikään <i>asuntomme ei ollut kovin iso</i> , ja minä tarvitsin ympärilleni
3	etenkään näkynyt. <i>Arttu ei vaikuttanut kovin pettyneeltä</i> , ja minä sanoin että min
4	aisemille. Rajalla <i>jono ei näyttänyt kovin pitkältä</i> , mutta se ei tuntunut liikku
5	Ä ei ollut, vaikea, sillä <i>kaupunki ei ole kovin suuri</i> eikä siellä ole <i>kovin</i> monta nak
6	myös hymähtää mutta <i>yritys ei ollut kovinkaan onnistunut</i> . Hän henkäisi syvään,
7	ttöjen anatomiaa. <i>Kasvot eivät olleet kovinkaan tärkeitä</i> , mutta haju oli. Naiset eivät
8	Imeksi mainittua <i>vaihtoehtoa en pidä kovinkaan todennäköisenä</i> . Asserin ei yleensä
9	lainen valitti, että <i>työpäivä oli jäänyt kovin lyhyeksi</i> , Hartikainen peitteli mootto

Jos edellisissä tapauksissa pystyi kolmijäsenisten klustereiden perusteella rakentamaan selvän, muista rakenteista frekvenssinsä perusteella erottuvan klusterin, *oikein*-määritteen kohdalla tämä ei onnistu aivan yhtä helposti: määritteen erilaisten klustereiden esiintymistäajuuksissa ei ole yhtä selvää eroa kuin sen synonyymeilla. Kaksi positiota osoittautuu kuitenkin selvästi tietyille kolligaaiteille varatuiksi, nimittäin V1-asema verbeille ja O1 etupäässä adjektiiveille. Taulukon 6.19 perusteella yleisimmäksi viisijäseniseksi klusteriksi voisi abstrahoida rakenteen

<PRO VER *oikein* ADJ LL>.

Tämä rakenne ei kuitenkaan osoittaudu yleiseksi aineistossa, sillä se esiintyy vain viisi kertaa (esiintymät konkordanssissa 6.17 riveillä 1–5). Toinen yhtä yleinen rakenne on <PRO VER *oikein* ADV LL> (rivit 6–10), ja neljä kertaa esiintyviä syntagmoja on kolme erilaista. Tavallisimmaksi klusteriksi sen sijaan paljastuu rakenne

<SUB VER *oikein* ADJ SUB>,

joka sekin on aineistossa kuitenkin vain kuusi kertaa (5,0 % esiintymistä, rivit 11–16). *oikein*-määritteen syntagmoista ei siis näytä erottuvan selvästi yhtä yleistä rakennetta, kuten sen kahden synonyymien syntagmoista. Tämä astemääräte näyttää siten olevan syntagmaattiselta rakenteeltaan tasaisempi, koska se ei suosi tiettyä rakennetta. On kuitenkin muistettava, että tästä määritteestä on vähiten esimerkkitaupauksia korpuksessa. Lisäksi konkordanssin 6.17 rivien 6–10 tapaukset ovat kaikki peräisin kahdesta teoksesta eli edustavat vain kahden kielentuottajan idiolektia. Tulosta voi pitää *oikein*-määritteen osalta siten enemmänkin suuntaa antavana.

Taulukko 6.19. *oikein*-määritteen kolligaattiklusterit.

	V2	V1	noodi	O1	O2	f
1	PRO	VER	<i>oikein</i>			17
2	SUB	VER	<i>oikein</i>			15
3	LA	VER	<i>oikein</i>			12
4		VER	<i>oikein</i>	ADJ		35
5		VER	<i>oikein</i>	ADV		26
6		LA	<i>oikein</i>	ADJ		11
7			<i>oikein</i>	ADJ	SUB	31
8			<i>oikein</i>	ADV	LL	26
9			<i>oikein</i>	ADJ	LL	22

Vasemmalle, positioon V3 jatkuva klusterirakenne on *oikein*-määritteelle mahdollonta rakentaa, sillä kolligaattien variaatio on varsin kirjavaa tässä positiossa. Konkordanssista 6.17 näkyy kuitenkin, että lauseenalkuisuus on verrattain yleinen positiossa V3 (rivit 1–7 ja 12–13) ennen kursivoituja klustereita. Lisäksi kolligaattianalyysissä tuli ilmi, että lauseenalkuisuus on V3-kolligaateista yleisin 24,8 prosentin osuudellaan. Voidaan siten pitää todennäköisenä, että mikäli aineistoa kasvatettaisiin, lauseenalkuisuus nousisi kuusijäsenisen klusterin positioon V3 ja samalla luultavasti paljastuisi tyypillisin klusterirakenne, joka ei tässä analyysissä erottunut.

Konkordanssi 6.17. Esimerkkejä *oikein*-astemääritteen kolligaattiklustereista.

1	Otin yhteyttä fysioterapeuttiin. <i>Hän oli oikein viehättävä</i> , ja nimeltään Taina. Hänellä
2	katsomaan nakkioskioskin paikkaa. <i>Se oli oikein hyvä</i> . Tontti oli pieni jättömaan palan
3	ut päästä lähtemään. - Ei tarvitse, <i>se on oikein kaunis</i> . Ajattelin vain että... oliko se E
4	...tuota? Mannerheim: - Kyllä. <i>Tämä on oikein hyvää</i> . Katsokaa nyt. Hmm... herkul
5	dramaattisesti ovensuusta. Ei, <i>tämä on oikein mainiota</i> , minä sanoisin korvat heilue
6	taituroi ne jaloilleen. <i>Hän muisti oikein hyvin</i> milloin ja mistä ja millä hinnall
7	Kummisetä ravisti päätään. - <i>Minä voin oikein hyvin</i> , hän sanoi pontevasti, -
8	ja ilmiantaminen. Sen asian <i>minä tiedän oikein hyvin</i> . Sitten Helenan jalkaa alkoi
9	sinua siinä. Tämän asian <i>minä osaan oikein hyvin</i> . Eivät ne mitään opi mutta ne
10	tanssijan vartalo. Ja tämän <i>minä muistan oikein hyvin</i> . Katson Jelenaa ja sanon itsöks
11	Sariin, ja Sari sanoi <i>Topin olevan oikein hauska mies</i> , suorastaan huvittava jos
12	ylöspäin reisi jäi paljaaksi. <i>Rouvalla oli oikein hurmaavat jalat</i> . Lopuksi sipaisin
13	ei tapeta äijäraahujen takia. <i>Lolita oli oikein hyvä lehmä</i> , kiltti ja sävysä. Se lypsi
14	in vähänkin pyytää. Ja sitten <i>Helena sai oikein ison lohen</i> . Veijo oli soutamassa ja
15	Se selitti innoissann isäle, että <i>Juhola on oikein soma hellu</i> , joku yläperän lappala
16	ne näkisi hyvin. Sarin mielestä <i>aitta oli oikein viehättävä paikka</i> kynttilänvalossa.

Kolligaattiklustereiden tarkastelu osoittaa, että astemääritteiden yleisimmät viisi- ja kuusijäseniset kolligaattisyntagmat eivät kata kovin suurta osuutta syntagmavai-koimasta, koska niiden osuudet jäävät kaikilla määritteillä kymmeneen prosenttiin tai selvästi sen alle. Lyhyemmät klusterirakenteet kattavat sen sijaan luonnollisesti suuremman osan esiintymistä: yleisimpien klusterirakenteiden osuudet vaihtelevat *oikein*-määritteen 28,9 prosentista (<VER *oikein* ADJ>, n = 35) *kovin*-määritteen 35,8 prosenttiin (<VER *kovin* ADJ>, n = 63).

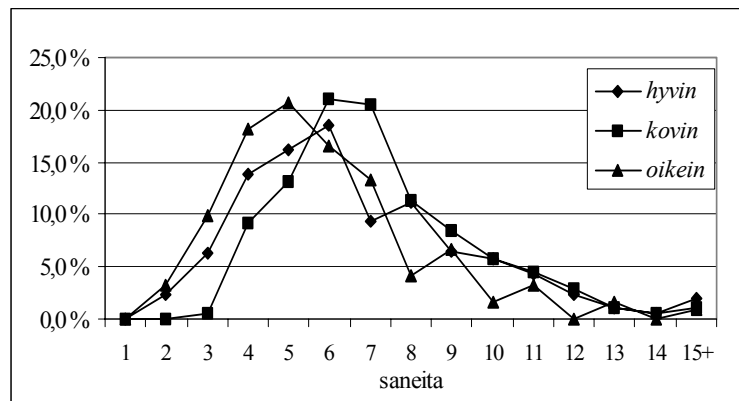
Kun edellä kuvatut koko tarkastelualueen kattavat klusterirakenteet esitetään allekkain, osoittautuu, että lauseasemakolligaatti on rakenteissa mukana vaihtelevasti: *hyvin*-määritteellä se on klusterin molemmissa ääripäissä, *kovin*-määritteellä oikeassa kontekstissa ja *oikein*-määritteellä mahdollisesti molemmissa.

hyvin <LA PRO VER *hyvin* ADJ LL>
kovin <SUB VER VER *kovin* ADJ LL>
oikein <?LA SUB VER *oikein* ADJ SUB/(LL)>

Klusterianalyysi paljastaa *oikein*-määritteen kuitenkin synonyymeistaan erottuvaksi, sillä sillä on muista poiketen myös sellaisia kolmiosisia klustereita, joissa lauseenalkuisuus on yhtenä kolligaattina vasemmanpuoleisessa kontekstissa positioissa V2 ja V1 eli suhteellisen vakiintuneena osana *oikein*-määritteen kontekstuaalista valintaa. Edelliset klusteritaulukot ja yllä olevat kuusiosaiset klusterirakenteet osoittavat, että lauseasemakolligaatti on eri tavoin mukana määritteiden syntagmoissa. Tämän perusteella voidaankin olettaa, että määritteet esiintyvät eripituisissa lauseissa. Viimeinen kolligaattien klusteroitumiseen liittyvä seikka, jota tässä käsitellään, onkin astemääritteiden kontekstuaalinen lausepituus.

6.2.4. Lausepituus synonyymeja erottamassa

Klusterirakenteiden perusteella suurin lausepituus näyttäisi olevan *kovin*-määritteellä, jolla ei ole yleisimmässä kuusiosaisessa klusterissa, eikä myöskään kolmiosaisissa, mukana lauseenalkuisuutta, toisin kuin sen synonyymeilla. *hyvin*- ja *oikein*-määritteiden välille on vaikea tehdä eroa, koska jälkimmäisen klusterirakenteet ovat määritteen esiintymien vähäisyyden vuoksi osittain hypoteettisia. Lausepituuden tarkistamiseksi on aineiston astemääritteiden kotekstista laskettu määritteen sisältävien lauseiden sanemäärät. Tulokset on esitetty graafisesti kuviossa 6.4. Eripituisten lauseiden osuudet osoittautuvat jossain määrin erilaisiksi määritteittäin. Lyhyiden, kaksi- ja kolmisanaisten lauseiden osuus on *kovin*-määritteellä pienempi kuin sen synonyymeilla; pitempien yli kymmensanaisten lauseiden osuuksien määrissä ei ole selvää eroa, joskin *oikein* poikkeaa *kovin*- ja *hyvin*-määritteiden linjasta hieman, mutta kyse on kuitenkin yksittäistapauksista.



Kuvio 6.4. Kotekstilauseiden sanemäärät.

Lauseiden tyypillisin sanemäärä (moodi) osoittautuu synonyymeilla hieman erilaiseksi: *hyvin*- ja *kovin*-määritteillä se on 6 sanetta ($n = 65$ ja $n = 37$) ja *oikein*-määritteellä 5 sanetta ($n = 25$). Aineiston perusteella *oikein* on siis tyypillisimmin viisisanaisten lauseissa eli näistä kolmesta määritteestä lyhimmissä, ja *kovin* ja *hyvin* taas tätä pitemmissä kuusisanaisten lauseissa. Lauseiden sanemäärien keskiarvo ja hajonta näyttävät taulukon 6.20 mukaisilta; taulukkoon on myös merkitty keskiarvojen eron tilastollinen merkitsevyys. Lauseiden sanemäärien keskiarvo tukee osittain moodianalyysia, sillä keskiarvot vaihtelevat määritteittäin ja jälleen *oikein* näyttää esiintyvän lyhimmissä lauseissa; ero on tilastollisesti merkitsevä suhteessa molempiin muihin määritteisiin. Testi erottaa nyt kuitenkin myös kaksi muuta määritettä toisistaan, sillä niiden lausepituuksien keskiarvo osoittautuu selvästi erilaiseksi ja tilastollisesti merkitseväksi.

Taulukko 6.20. Lauseiden sanemäärien keskiarvot ja hajonta sekä keskiarvojen erojen tilastollinen merkitsevyys astemääritteittäin.

	<i>hyvin</i>	<i>kovin</i>	<i>oikein</i>
keskiarvo (saneita)	6,68	7,28	5,83
vaihteluväli (saneita)	14 [2–16]	19 [3–21]	13 [2–15]
	<i>hyvin–kovin</i>	<i>hyvin–oikein</i>	<i>kovin–oikein</i>
merkitsevyys	$z = 2,40$ ($p < 0,05$)	$z = 3,19$ ($p < 0,01$)	$z = 4,87$ ($p < 0,001$)

Näin selviää, että *kovin* on synonyymeihinsa verrattuna pisimmissä lauseissa, *oikein* lyhyimmissä ja *hyvin* jää näiden välille. *oikein*- ja *hyvin*-määritteiden kotekstit ovat lauseiden sanemäärien jakauman suhteen vähemmän varioivia kuin *kovin*-määritteen kotekstin lauseet. Tämä näkyy saneiden vaihteluvälin pituudesta, joka on kahdella ensin mainitulla selvästi pienempi kuin *kovin*-määritteellä. Viimeksi mainitulla vaihteluvälin pituutta kasvattaa huomattavasti muita pitempi 21-saneinen lause (esimerkki konkordanssissa 6.18 rivillä 9). Kun nämä poikkeavan pitkät lauseet poistetaan, jäävät astemääriteparien erot kuitenkin edelleen merkitseviksi⁶⁰. Ilman kahta epätyypillisen pitkää lausettakin *kovin* esiintyy synonyymejaan pitemmissä lauseissa. Näin ollen astemääritteiden kotekstin tyypillisintä lausepituutta osoittavat moodit ja lausepituuden keskiarvot korostavat astemääritteiden eroja.

Yksi selitys *kovin*-määritteen lausepituuteen löytyy kieltoverbistä EI sen kotekstissa. Kuten jo aiemminkin on todettu, kieltoverbi pidentää lauserakenteita kasvatamalla osaltaan verbirakenteiden pituutta. Kieltoverbi ei kuitenkaan näyttäisi yksinään selittävän lausepituuksien eroa, sillä *oikein*- ja *kovin*-määritteiden keskiarvojen erotus on lähes puolitoista sanetta. Tämä tuntuu viittaavan siihen, että kotekstista on löydettävissä muitakin tekijöitä, jotka pidentävät *kovin*-määritteen kotekstilauseita ja vastaavasti pienentävät muiden, erityisesti *oikein*-määritteen kotekstilauseita. Lauseiden pituuteen näyttää vaikuttavan myös se, millaisissa lauseissa määritteet esiintyvät. Kun tarkastellaan sekä lyhyimpien 1–3 sanetta että pisimpien, yli 10 saneetta sisältävien lauseiden osuutta, havaitaan näiden määrät selvästi erilaisiksi. Ensinnäkin lyhyiden lauseiden osuudet eroavat selvästi: *kovin*-määritteellä niitä on 0,6 % kaikista lauseista, *hyvin*-määritteellä 8,5 % ja *oikein*-määritteellä 13,2 %. Pitkien lauseiden osuudet ovat vastaavasti *kovin*-määritteellä ja *hyvin*-määritteellä 15,9 % ja *oikein*-määritteellä noin puolet vähemmän, 7,4 %. Erityisesti *oikein*-määritteen, mutta myös *hyvin*-määritteen lyhyet lauseet, joiksi tässä on laskettu myös finiittiverbittömät elliptiset rakenteet, ovat usein osa vuoropuhelua (konkordanssissa 6.18 rivit 1–3) tai lyhyttä (elliptistä) kuvausta ja kommentointia (rivit 4 ja 5) ja erityisesti *hyvin*-määritteen yhteydessä lyhyitä tilalauseita (rivi 6). *kovin*-määritettä ei taas näy käytettävän tällaisissa rakenteissa. Pitkät lauseet ovat tyypillisimmin polveilevia, pitkiä määriteketjuja tai rinnasteisia konstituentteja sisältäviä rakenteita, joissa kuvataan kerronnan kohteena olevia tapahtumia, asioita ja henkilöitä (rivit 7–9). Nämä ovat vuorostaan tavallisempia *hyvin*- ja *kovin*-määritteellä mutta harvinaisia *oikein*-määritteellä.

⁶⁰ *hyvin* vs. *kovin* $z = 1,974014$ ($p < 0,05$); *kovin* vs. *oikein* $z = 4,690112$ ($p < 0,001$).

Konkordanssi 6.18. Esimerkkejä lyhyistä ja pitkistä lauseista astemääritteiden koteksissa.

1	”, vastasi toinen, rauhallisempi ääni. ” <i>Oikein pettynyt!</i> ” Louner kurkisti ovilevyn ja	
2	syömään?” Gimma tekee tarjouksen. ” <i>Oikein kivaan paikkaan.</i> Paljonks oot valmis i	
3	”Millainen olo sinulla on?” kysyin. ” <i>Hyvin bysanttilainen.</i> ” ”Oliko Bysantti sama k	
4	me halpaa marokkolaista punaviiniä. <i>Hyvin punaista.</i> Hiukan lämmintä. Sen maku t	
5	unsaasti mureaa sianlihaa. <i>Maistuivat oikein hyvältä.</i> Sikalan hoitaja ohjasi minut e	
6	, sillä satoi räntää ja meitä paleli. <i>Oli hyvin hiljaista.</i> En muista nähneeni kaupunki	
7	n, ohimoita jomotti. <i>Niinpä hän hyvin nopeasti, kivuttomasti ja äänettömästi.</i> <i>päätti tappaa miehen</i> Sitten hän halusi tehdä jotakin aivan m	
8	<i>Flanderilainen mysteerinäytelmän hyvin suosittua Elckerlijc-mysteeriä tuomio-</i> <i>esittäjäseurue oli saanut Aschaffenburgin tuomioherran kapitulista kirkon aukiolla, portailta, siirtyä siitä</i> <i>luvan näytellä nykyisin</i> eteishalliin ja tympanonin alitse kirkon pääkäytävälle.	
9	kymmenen uutiset pizzojaan tyyty- <i>kovin tuttua miestä.</i> - Voi hyvä Jumala! Irene väisinä mussutellen, <i>kunnes niiden parahti kauhuissaan.</i> <i>lopuksi heidän silmiensä eteen le-</i> <i>vähti Onnelan sauna ja kolme kyl-</i> <i>pytakkeihin pukeutumutta, kiih-</i> <i>tyneesti ja pateettisesti puhuvaa ja</i>	

6.2.5. Synonyymien *hyvin*, *kovin* ja *oikein* kolligatiivinen profiili

Tutkittujen astemääritteiden kolligatiivinen kuva on esitetty kootusti taulukossa 6.21, johon on kerätty tyypillisimmät kolligaatit koko tarkastelualueelta ja positiivittain, kolligaattien klusteroituminen ja lausepituus. Kolligaattianalyysi osoitti, että määritteet eroavat toisistaan kolligaattien osuuksien perusteella, positiivittain tarkasteltuna jonkin verran tyypillisimmän kolligaatin mukaan, klusteroitumisen perusteella ja viimein lausepituuden mukaan. Viimeiseksi luodaan lyhyt katsaus siihen, miten määriteparit eroavat toisistaan synonyymisyyden suhteen. Analyysin perusteella tuli ensinnäkin esiin se, että kolligaattien kirjo on samanlainen kaikilla astemääritteillä kaikissa tarkastelluissa asemissa (V3–O2). Tämä korostaa astemääritteiden synonyymisyyttä. Sen sijaan tilastolliseen merkitsevyyteen perustuva kolligaatioiden *osuuksiin* keskittyvä tarkastelu toi esiin eroja astemääritteiden välille, mikä taas korostaa sitä, että nämä synonyymiset astemääritteet eroavat selvästi kontekstuaalisilta piirteiltään toisistaan.

Taulukko 6.21. Astemääritteiden *hyvin*, *kovin* ja *oikein* kolligationaalinen profiili.

	<i>hyvin</i>	<i>kovin</i>	<i>oikein</i>
Kolligaatit			
frekventein alueella V3–O2	VER 18,4 %	VER 30,0 %	VER 19,2 %
frekventein positioittain			
V3	LA 24,1 %	SUB 29,0 %	LA 24,8 %
V2	VER 27,0 %	VER 53,4 %	SUB 28,1 %
V1	VER 42,9 %	VER 58,5 %	VER 50,4 %
O1	ADJ 66,5 %	ADJ 60,2 %	ADJ 55,4 %
O2	LL 52,3 %	LL 40,9 %	LL 42,1 %
Klusterit			
tyypillisin			
3-jäseninen	< <i>hyvin</i> ADJ LL> 33,2 %	<VER <i>kovin</i> ADJ> 35,8 %	<VER <i>oikein</i> ADJ> 28,9 %
6-jäseninen	<LA PRO VER <i>hyvin</i> ADJ LL> 4,3 %	<SUB VER VER <i>kovin</i> ADJ LL> 5,1 %	<?LA SUB VER <i>oikein</i> ADJ SUB> 5,0 %
Lausepituus (saneita)	6,68	7,28	5,83

Taulukkoon 6.22 on koottu kaikki astemääritteiden kontekstissa esiintyvät, kvantitatiivisin perustein astemääritteet toisistaan erottavat kolligaatit. Kuten näkyy, erottavien muuttujien määrä vaihtelee positioittain ja astemääritepareittain. Kolligaattien osuus erottaa astemääritteet toisistaan useimmin V2-asemassa, ja positioissa V3 ja O2 eli kauimpana hakusanoista erottavia muuttujia on vähiten. Tämä kuvastaa sitä, että erottavat tekijät löytyvät läheltä hakusanoja. Kolligaationalyysin tuottama tulos on siten rinnasteinen kollokaatioanalyysin tulokseen, joka myös osoitti, että synonymiaa rikkovat tekijät löytyvät selvimmän nooidien lähispositioista (ks. aiemmin taulukkoa 6.7 s.98). Vasemmanpuoleinen konteksti eroaa oikeanpuoleisesta siinä, että vasemmassa kontekstissa on lähisynonymisyyttä korostavien kolligationaalisten erojen määrä suurempi. Kun astemääriteparien kolligationaalisia eroja tarkastelee yllä olevan koonnin perusteella, näyttää siltä, että *hyvin*–*kovin*-parin jäsenet eroavat kolligaatiosuhteiltaan eniten toisistaan, sillä niitä erottavat peräti yhdeksän kolligaatin osuuksissa ilmenevät erot ja kolligaattien suhde-eroja on jokaisessa tarkastelualueen asemassa. Kahden muun parin välillä ero on pienempi: *kovin*–*oikein*-paria erottavat seitsemän kolligaatin osuuksien erot ja *hyvin*–*oikein*-paria kuuden kolligaatin osuudet, eli nämä astemääritteet näyttäisivät olevan pareista synonyymisimpiä.

Taulukko 6.22. Kolligaatit, joiden osuudet erottavat astemääritteet toisistaan tilastollisin perustein.

Positio	<i>hyvin - kovin</i>	<i>hyvin - oikein</i>	<i>kovin - oikein</i>
V3	LA	–	LA
V2	VER SUB PRO	VER LA	VER SUB LA
V1	VER KON	LA	LA
O1	KVA	ADJ ADV	ADV KVA
O2	VER LL	VER	–

Astemääritteiden kontekstista erottuu selkeästi kaksi kolligaattia, jotka ovat lähisyntyisyyttä korostavia: lauseenasemakolligaatti ja verbikolligaatti. Lauseenasemakolligaatti, erityisesti lauseenalkuisuus, erottaa astemääritteet toisistaan seitsemästi – kaikkiaan neljässä eri positiossa ja kaikissa astemäärittepareissa. Verbikolligaatti taas kuudesti, ja myös se erottaa kaikkien parien määritteitä mutta vain kolmessa eri asemassa. Muiden kolligaattien merkitys lähisyntyisyyttä korostavana muuttujana ei ole yhtä selkeä, mikä tarkoittaa, että erityisesti verbien ja lauseenaseman osuuksien erot erottelevat määritteitä ja korostavat nimenomaan niiden lähisyntyisyyttä.

Myös kolligaattien klusteroituminen määritteiden lähikontekstissa paljasti määritteiden kontekstuaaliset merkitykset hieman erilaisiksi. Lisäksi kolligaattien ja erityisesti niiden klusteroitumisen analysoiminen johti synonyymisten sanojen kontekstin analyysia uusille väylille, nimittäin lausepituuden merkityksen havaitsemiseen yhtenä mahdollisena synonymiaa rikkovana tekijänä. Lauseen pituus lienee ainakin osittain sidoksissa sen sisältämän ilmauksen funktioon: esimerkiksi tilalauseet ja vuoropuhelua kuvastavat ilmaukset lienevät herkemmin lyhyempiä kuin tapahtumia ja ihmisiä kuvailemaan pyrkivät ilmaukset, vaikkakin tässä esiintyneen runsaasti kirjoittaja- ja tyylikohtaista vaihtelua. Näin ollen itse ilmauksen funktiota ja mahdollisesti myös lausetyyppejä (esim. luonnehtivien predikaatiivilauseiden voi olettaa tyypillisesti sisältävän astemääritteitä) voitaisiin pitää yhtenä kolligatiivisena ominaisuutena, jota lienee syytä tarkastella, kun pyritään kuvaamaan leksikaalisen yksikön kontekstuaalisia kytkentöjä.

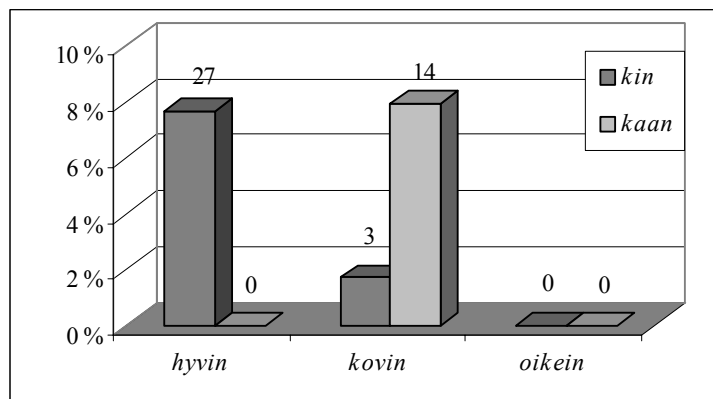
6.3. Morfologinen preferenssi

Tässä jaksossa siirrän tarkastelun morfologiseen preferenssiin, jota olen ehdottanut yhdeksi, erityisesti synteettisten kielten sanaston kontekstuaaliseen merkitykseen kuuluvaksi piirteeksi. Aikaisemmassa tutkimuksessani (Jantunen 2001b), jossa käsitelin noodien morfologista preferointia, tutkimuskohteena olleet lekseemit poikkesivat syntaktisilta funktioiltaan jonkin verran, mikä puolestaan selitti niiden erilaisen muoto-opillisen preferenssin. Onkin haasteellista testata, voiko morfologista preferenssiä olla löydettävissä sellaisten sanojen kesken, joilla ei ole variaatiota syntaktisessa funktiossa – kaikki tarkasteltavat sanathan ovat tässä astemääritteen funktiossa.

Edellä luvussa 2.3.5 olen jaotellut morfologisen preferenssin kontekstuaaliseen ja paradigmaattiseen preferenssin, joista edellinen käsittää sekä lemman eri sanamuotojen kontekstuaalisen preferoinnin että koko lemman suhteen kontekstin eri lekseemien sanamuotoihin. Astemääritteen tapauksessa ensisijaiseksi tarkastelu-kohteeksi muodostuu lemman suhde kontekstissa olevien lekseemien eri taivutusmuotoihin, koska astemääritteet ovat taipumattomia, ja siten niillä ei voi esiintyä paradigmaattista variaatiota, joka on edellytys kahdelle muulle morfologisen preferenssin muodolle. Silti astemääritteet eivät ole täysin vailla suffiksointia, mikä on näkynyt jo joissakin aiemmissa aineistoesimerkeissä: ainakin tässä käsitellyt astemääritteet näyttävät saavan jossakin määrin suffiksoidun liitepartikkelin; niinpä esimerkeissä on ollut muotoja *hyvinkin* (*hyvin* + *kin*) ja *kovinkaan* (*kovin* + *kaan*). Liitepartikkelin liittämistä ei voida kuitenkaan yksiselitteisesti lukea taivutuksen piiriin, koska prosessi ei ilmaise suffiksoidun ilmauksen syntaktis-funktionaalisia suhteita lauseessa vaan pikemminkin lausetason prosodisluonteista, pragmaattista täydennystä (Karlsson 1982: 175, 225) ja tekstuaalisia sekä emotionaalisia tehtäviä (Hakulinen ja Karlsson 1979: 327). Karlsson (mts. 225) käyttääkin liitepartikkelin suffiksoinnista käsitettä *ulkoliitäntä* korostaakseen suffiksoinnin erilaisuutta suhteessa varsinaiseen taivutukseen. Toisaalta on kuitenkin huomattava, että abstraktioihin *lekseemi* ja *lemma* sisällytetään myös liitepartikkelilliset muodot, jolloin ne ovat ainakin jossain määrin osa sanan paradigmaa. Lisäksi suomen liitepartikkelit eivät ole itsenäisiä morfeemeja vaan aina liittyneitä johonkin sanamuotoon, joten niitä ei myöskään tavata kontekstuaalisena irrallisena aineksena kontekstista – eikä niitä siten voi tarkasti ottaen pitää kontekstuaalisina yksikköinä. Morfologisen preferenssin tarkastelussa lasken siten astemääritteiden liitepartikkelilliset muodot paradigmaattiseen preferenssiin kuuluvaksi. Näin ollen *hyvin*-, *kovin*- ja *oikein*-määritteiden kontekstuaalisten merkitysyksikköjen kuvausta täydennetään sekä paradigmaattiseen että kontekstuaaliseen morfologiseen preferenssiin kuuluvilla ominaisuuksilla.

6.3.1. Paradigmaattinen morfologinen preferenssi

Liitepartikkelin sisältävät muodot on saatu aineistosta käyttämällä komentojonoja *hyvin**, *kovin** ja *oikein**. Suomen kielen liitepartikkeleista (*hAn*, *kAAAn*, *kin*, *kO*, *pA*, *kA*, *s*) astemääritteihin *hyvin*, *kovin* ja *oikein* liittyneinä esiintyvät vain polaariset liitepartikkelit *kin* ja *kaan*. Astemäärite + *kin/kaan* -rakenteet eivät muodosta kiinteää leksikaalistunutta ilmausta, jollaisia ovat muun muassa *arvatenkin* ja *tietenkin* (leksikaalistuneista muodoista ks. Mäkelä 1993: 20–23). Yhtymä *hyvin* + *kin* voi kuitenkin muodostaa leksikaalistuneen ilmauksen merkityksensä myönnyttelevä ’kyllä(kin), tietenkin, sittenkin, sentään, ainakin, ehkäpä’, kuten lauseessa *Jospa meitä hyvinkin onnistaisi* (CD-PS s.v. *hyvin*), mutta näissä tapauksissa *hyvin* ei ole astemääritteen funktiossa⁶¹. Näistä kahdesta liitepartikkelista *kin* esiintyy pääasiassa myönteisissä ja *kaan* kielteisissä yhteyksissä (Sadeniemi 1942: 285; Hakulinen ja Karlsson 1979: 269; ks. myös Mäkelä mts. 2), joten on oletettavissa, että *kin*- ja *kaan*-partikkelien esiintyminen noudattaa astemääritteiden todettua kielto- ja myöntöhakuisuutta (taulukko 6.6 s. 92). Aineistosta tavattujen liitepartikkelillisten tapausten määrät on kuvattu astemääritteittäin kuviossa 6.5.



Kuvio 6.5. Liitepartikkelien suffiksoituminen määritteisiin.

Kuvion valossa astemääritteiden paradigmaattinen morfologinen preferenssi osoitetaan varsin erilaiseksi: *hyvin*-määritteen suffiksoituu ainoastaan *kin* ($n = 27$), *kovin*-määritteen tyypillisimmin *kaan* ($n = 14$), mutta *oikein* ei saa minkäänlaista liitepartikkelia. Liitepartikkelien suffiksoituminen näyttää noudattavan astemääritteiden kieltohakuisuutta, sillä *kin* on tyypillisesti myöntöhakuisen *hyvin*-määritteen liitteenä ja *kaan* kieltohakuisen *kovin*-määritteen. Täydennysjakauma ei ole kuitenkaan täysin ehdoton – kuten ei näillä liitepartikkeleilla yleensääkään (ks. Hakulinen ja Karlsson mts. 329–330) – sillä *kovin* saa liitteeseen satunnaisesti myös *kin*-liit-

⁶¹ *hyvinkin* ja *kovinkin* ovat kuitenkin jossain määrin leksikaalistuneet, sillä niiden merkitys ei ole ’myös + määrän ilmaiseminen’ vaan ilmauksista tulee jossain määrin tähdentäviä ellei jopa yllätyksellisyttäkin ilmaisevia.

teen. Esimerkkejä liitepartikkelien suffiksoitumisesta määritteisiin seuraavassa konkordanssissa.

Konkordanssi 6.19. Esimerkkejä liitepartikkelillisista astemääritteistä.

1	uksia. Me luomme skenaarioita, ja	<i>hyvinkin</i>	eripituisille aikajänteille. Muutamasta
2	taiset tiedot alkoivat tuntua Askosta	<i>hyvinkin</i>	tärkeiltä ja merkityksellisiltä. Pari päivä
3	neet keskenään, oppivat tuntemaan	<i>hyvinkin</i>	tarkkaan toistensa elämän ja ajatukset.
4	Olisin joutunu sokkiin ja vuotanu	<i>hyvinki</i>	pitkään. Aattelin että nyt Kolkko on ko
5	n läpäisevä. Lasta voidaan kohdella	<i>hyvinkin</i>	julmasti, koska tällä oletetaan olevan ed
6	olimerkitys, joka saattaa muuttua	<i>hyvinkin</i>	konkreettiseksi. Niin kuin nyt punainen
7	mitään. Mutta mie näin, että äiti oli	<i>kovinki</i>	kiihyksisään. Punaset läikät palo sen po
8	ttä jo vain veikkoset. Se koitti olla	<i>kovinkin</i>	pontevana siinä, sekä suorana silitteli
9	kaksi kalaa kuivalla maalla, muka	<i>kovinkin</i>	erilaisia, toinen väittää olevansa tappaja
10	iti ei antanut minun istua selässään	<i>kovinkaan</i>	kauan. Hän valitti painoni sattuvan.
11	sta parrasta huolimatta hän ei ollut	<i>kovinkaan</i>	vanha mies. Sisällä aulassa meitä odotti
12	yös hymähtää mutta yritys ei ollut	<i>kovinkaan</i>	onnistunut. Hän henkäisi syvään, levitti

6.3.2. Kontekstuaalinen morfologinen preferenssi

Astemääritteiden kontekstin muoto-opilliseen koodaukseen käytetty Conexorin Machines Phrase Tagger -ohjelman morfologisen analyysin tulokset on esitetty tiivistetysti taulukossa 6.23. Taulukkoon on koottu jälleen yleistettävyyperiaatteen mukaisesti vain tyypillisimmät kontekstissa esiintyvien kollokaattien morfologista rakennetta kuvaavat merkkijonot. Raja-arvona on pidetty tunnisterypään esiintymistä vähintään viisi kertaa sataa astemäärite-esiintymää kohti. Taulukossa on ilmoitettu morfologista tulkintaa kuvaava merkkijono ja sen suhteellinen frekvenssi. Kaikille yhteiset morfologisessa kuvauksessa olevat merkkijonot muodostavat synonymiaa korostavan muoto-opillisen kontekstin, joka on (ainakin näille kolmelle) astemääritteelle tyypillinen kaunokirjallisuudessa. Tähän kuuluvat substantiivin yksikön nominatiivi, genetiivi ja partitiivi, adjektiivin yksikön nominatiivi ja partitiivi, pronominin yksikön nominatiivi sekä verbin aktiivin indikatiivin imperfektin yksikön kolmas persoona ja aktiivin ensimmäinen infinitiivi. Näistä muodoista on esimerkkejä tässä listatussa järjestyksessä konkordanssissa 6.20 (s. 131) riveillä 1–16.

Kaikkien edellä mainittujen tunnistejonojen osuudet vaihtelevat jossain määrin, mutta erityisesti vaihtelua on tunnistejonojen <N SG NOM>, <A SG NOM> ja <V ACT IND PAST SG3> osuuksissa: ensimmäisen osuus on *kovin*-määritteellä vain 25,6 sataa esiintymää kohti, kun *hyvin*-määritteellä se on 39,5 ja *oikein*-määritteellä peräti 48,8. Myös <A SG NOM>:n osuus on *kovin*-määritteellä muita pienempi (21,6 vs. 36,1 ja 34,7), ja samoin kolmannenkin tunnistejonon, <V ACT IND PAST SG3>:n osuus, joka on vain 16,5, kun muilla se on noin kaksinkertainen. Nominien

Taulukko 6.23. Astemääritteiden *hyvin*, *kovin* ja *oikein* kontekstin morfologisen analyysin tulokset. (Lihavointi = kaikille astemääritteille yhteinen koodaus, ** = vain kyseisen astemääritteen saama koodaus.)

<i>hyvin</i>		<i>kovin</i>		<i>oikein</i>	
tunnisteet	f %	tunnisteet	f %	tunnisteet	f %
N SG NOM	39,5	**V ACT SG3	26,1	N SG NOM	48,8
V ACT IND PAST SG3	36,9	N SG NOM	25,6	A SG NOM	34,7
A SG NOM	36,1	**V ACT IND PAST SG NEG	21,6	V ACT IND PAST SG3	32,2
A SG PTV	15,6	A SG NOM	21,6	N SG GEN	20,7
N SG GEN	14,8	**V ACT PRES NEG	20,5	V ACT IND PRES SG3	19,0
V ACT IND PRES SG3	14,2	V ACT IND PAST SG3	16,5	PRON SG NOM	16,5
PRON SG NOM	10,5	N SG GEN	14,2	N SG PTV	14,9
A PL PTV	10,2	PRON SG NOM	13,1	V ACT INF1	11,6
**N PL NOM	9,7	A SG PTV	12,5	A SG PTV	9,9
N SG PTV	8,5	N SG PTV	11,4	**V ACT IND PRES SG1	8,3
PRON SG3 NOM	8,0	**V ACT SG1	10,2	N SG ADE	6,6
V ACT IND PAST SG1	8,0	V ACT INF1	9,7	V ACT IND PAST SG1	6,6
V ACT INF1	8,0	PRON SG3 NOM	8,0	A SG ABL	5,8
**N SG ELA	6,5	A PL PTV	8,0	A SG GEN	5,8
N PL PTV	6,3	**PRON SG PTV	7,4	**N SG INE	5,8
**V ACT IND PAST PL3	6,0	A SG ABL	5,7	PRON SG1 NOM	5,8
A SG GEN	5,7	**A SG TRA	5,1		
PRON SG1 NOM	5,7	N PL PTV	5,1		
V ACT PCP2 SG	5,4	V ACT PCP2 SG	5,1		
N SG ADE	5,1				

Tunnisteet: A = adjektiivi, N = substantiivi, PRON = pronomini, V = verbi; ADE = adessiivi, ABL = ablatiivi, ELA = elatiivi, GEN = genetiivi, INE = inessiivi, NOM = nominatiivi, PTV = partitiivi, TRA = translatiivi; SG = yksikkö, PL = monikko; ACT = aktiivi, IND = indikatiivi, PRES = presens, PAST = imperfekti, INF1 = 1. infinitiivi, PCP2 = 2. partisiippi

nominatiivimuotojen yleisyys muilla paitsi *kovin*-määritteellä näyttää ainakin osin selittyvän position O2 substantiivipredikaattien ja niitä määrittävien adjektiivien yleisyydellä näiden kontekstissa, onhan nominatiivi juuri predikatiivin tyypillinen sija ja identiteettipredikatiivin ainoa sija (Hakulinen ja Karlsson 1979: 194). Näistä käyvät esimerkiksi konkordanssin 6.20 rivit 1 ja 7. Lisäksi näiden astemääritteiden kontekstin samassa positiossa on runsaasti lauseenalkuisia yksikön nominatiivissa olevia substantiiveja, joita taas *kovin*-määritteellä ei ole montakaan. Rakenteen <V ACT IND PAST SG3> pieni määrä *kovin*-määritteen kontekstissa johtuu tämän astemääritteen kontekstin verbirakenteista: kieltoverbin runsaus selvästi vähentää myöntölauseille tyypillistä rakennetta.

Astemääritteillä on myös joitakin vain kullekin noodille tyypillisiä kontekstuaalisia tunnisterakenteita, jotka erottavat ne toisistaan (merkintä **). Erityisesti tämä koskee *kovin*-määritettä, jolla näitä on peräti kuusi. Tavallisimmin ne ovat verbimuotoja, joko kieltoverbin tai kieltomuotojen pääverbien taivutusmuotoja. Näin tämän astemääritteen kieltihakuisuus heijastuu myös sen kontekstin sananmuotoihin ja näkyy selkeänä morfologisena preferenssinä tiettyihin verbimuotoihin. Lisäksi *kovin* näyttää preferoivan jossain määrin myös pronomien yksikön partitiivia ja adjektiivin yksikön translatiivia, mutta näiden osuudet jäävät verbimuotoihin verrattuna pieniksi. *hyvin* näyttää preferoivan vuorostaan monikkomuotoja: sekä substantiivien monikon nominatiivi ja verbien monikon kolmas persoona ovat sen kontekstille tyypillisiä morfologisia muotoja. Vaikka niiden määrä ei ole yhtä suuri kuin *kovin*-määritteen verbimuotojen, ovat nämä muodot – etenkin yhdessä esiinty-

essään – tyypillisiä juuri *hyvin*-määritteelle, sillä muiden määritteiden kontekstissa nämä monikkomuodot ja erityisesti niiden yhteisesiintyminen ovat harvinaisia (esimerkkejä konkordanssin 6.20 riveillä 17–18).

oikein vuorostaan preferoi jossain määrin verbin preesensin yksikön ensimmäisiä muotoja, erityisesti näitä tavataan tapaa ilmoittavan *hyvin*-adverbin yhteydessä (rivit 19–20). Näistä seitsemästä tapauksesta kuitenkin viisi on peräisin yhdeltä tekstin tuottajalta, joten käsillä olevan aineiston valossa kysymyksessä ei liene laaja-alainen tendenssi. Myös *oikein*-määritteen <N SG INE> ja *hyvin*-määritteen <N SG ELA> saattavat olla osoitus astemäärittekohtaisesta morfologisesta preferoinnista, mutta ainakin tässä aineistossa näiden määrät ovat pienet ja siten niitä ei voi nostaa kovinkaan selviksi osoituksiksi preferoinnista.

Konkordanssi 6.20. Esimerkkejä morfologisia tunnistejonoja vastaavista kollokaateista (tummennettu).

N SG NOM		1	koskevaa mielipidetiedustelua. Se on	<i>hyvin</i>	merkittävä löytö , sillä tähän asti ei ole
		2	in pyytää. Ja sitten Helena sai oikein	<i>oikein</i>	ison lohen. Veijo oli soutamassa j
N SG GEN		3	i aineen kirjoittaminen veden alla ole	<i>kovin</i>	helppoa, etenkin kun veden lämpötila
		4	le otuksia. Yhenki äijän päästä lähti	<i>oikein</i>	pahan näkösiä olioita. Niillä oli pun
N SG PTV		5	amaan. Ehkä Regina auttoi. Kaksi	<i>kovin</i>	erilaista naista olivat ystäväystyneet. K
		6	rveltä alkoi nousta usvaa . - Minä olin	<i>hyvin</i>	hämmentynyt sen ensimmäisen tapaa
A SG NOM		7	inoa rakkaus, viimeisen rauhan kesän	<i>hyvin</i>	hellä rakastaja. Äiti jätti kertomatta et
		8	vosanoja upea tai suurenmoinen . Siis	<i>oikein</i>	ihailtavaa kaukonäköisyyttä. Voitot ti
A SG PTV		9	erehtyneistä nuorukaisista. Sehän oli	<i>kovin</i>	kirjallista , siitä kerrottiin mestariteoks
		10	lottuvalla näkemyksellä elävöitettyä.	<i>oikein</i>	hyvää illanjatkoa kaikille katsojillem
PRON SG NOM		11	Tunnen oloni syntisäkiksi, eikä se ole	<i>kovin</i>	kivaa, hän huokaisi. -- Omista tekemis
		12	llä oli tunne joka muistutti etäisesti	<i>hyvin</i>	kaukaista iltaa koulun joidenkin juhlie
V ACT IND PAST SG3		13	nkylän koulussa, niin minulle sattui	<i>oikein</i>	ikävä ja hankala oppilas, eräs Jari Ku
		14	masta halustaan. Kettusen Pekka oli	<i>hyvin</i>	kiinnostunut ihmisistä ja hän oli käyn
V ACT INFI		15	bolimerkitys, joka saattaa muuttua	<i>hyvinkin</i>	konkreettiseksi. Niin kuin nyt punain
		16	nneli tahtoo ujoa mallia. Silloin voit	<i>oikein</i>	hyvin mennä . - Oletko sinä aivan var
N PL NOM + V ACT IND PAST PL3		17	. Jostakin syystä lehdet tekivät kalat	<i>hyvin</i>	aroksi. Topi alkoi soutaa venettä taka
		18	in hyvin somaa. Puukengät kopisivat	<i>hyvinkin</i>	rytmikkäästi Askon takana. Paikallaa
V ACT IND PRES SG1		19	ilmiantaminen. Sen asian minä tiedän	<i>oikein</i>	hyvin. Sitten Helenan jalkaa alkoi sär
		20	minäkään ehdi. Minä viihdyn kotona	<i>oikein</i>	hyvin. Minä voin erinomaisesti, pare

Kaiken kaikkiaan astemääritteille *hyvin*, *kovin* ja *oikein* on osoitettavissa toisistaan poikkeava morfologinen preferenssi, ja se toteutuu sekä paradigmaattisena että kontekstuaalisena preferenssinä. Aineiston analyysissa tuli myös ilmi, että morfolo-

ginen preferenssi on tiukasti sidoksissa aiemmin kollokaatioanalyysissa esille tulleen kieltohakuisuuteen. Niinpä *kovin* saa kieltohakuisuutensa vuoksi kotekstiinsa tiettyjä negaatiota ilmentäviä muotoelementtejä, joko kieltoverbin taivutusmuotoja, liittomuotojen pääverbien kieltomuotoja tai kieltohakuisen *kaan*-liitepartikkelin. Myös kahden muun määritteen kotekstista on löydettävissä morfologista preferenssiä, mutta ne – etenkin kotekstuaalinen preferenssi – eivät ole toisistaan yhtä poikkeavia kuin *kovin*-määritteen osoittama preferenssi suhteessa synonyymeihinsa. Tässäkin tulee ilmi morfologisen preferenssin ja muiden kotekstuaalisten ominaisuuksien yhteys, sillä *kovin*-määritteen kotekstissa ilmenevä kieltohakuisuus, joka reaalistuu pääasiassa kieltoverbikollokaatteina, ohjaa vahvasti tämän astemääritteen morfologista preferenssiä ja erottaa määritteen synonyymeistaan. Vastaavasti on aiemmin toisessa yhteydessä (Jantunen 2001b; ks. myös lukua 4.2) tarkastelemieni adjektiivien TÄRKEÄ ja KESKEINEN morfologinen paradigmaattinen preferenssi selittynyt näiden adjektiivien syntaktisilla funktioilla: pääasiassa adjektiiviattribuuttina käytetty KESKEINEN noudattaa tyypillisesti nominatiivimuotoisen pääsanansa sijataivutusta, jolloin sen taajimmin esiintyväksi sijaksi erottuu nominatiivi. Sen synonyymien syntaktiset funktiot ovat taas moninaisemmat, mikä vuorostaan näkyy erilaisena morfologisena preferenssinä synonyymien välillä. Synonyymisia verbejä *mieltä* ja *pohtia* tutkinut Arppe (2002) onkin todennut, että pelkkä havaintoyksiköiden tarkastelu ei riitä selittämään niiden (tässä tapauksessa paradigmaattista) morfologista vaihtelua, vaan selitys löytyy kotekstin ja havaintoyksikön suhteesta:

Despite all my attempts to rule out the effect of word-external context, many of the inflectional features are interconnected with the lexical and syntactic context -- (mts. 31).

6.4. Semanttinen valinta

Viimeinen astemääritteiden kontekstuaalisia preferenssejä koskeva analyysi kohdistuu kotekstissa ilmeneviin semanttisiin piirteisiin. Esityksessä tarkastellaan astemääritteiden *hyvin*, *kovin* ja *oikein* kontekstien semanttista profiilia aloittamalla semanttisesta preferenssistä.

6.4.1. Semanttinen preferenssi

Havaintoyksiköiden semanttisia myötäesiintymiä koskevissa analyyseissa tarkastelun kohteena olevat ilmaukset ovat olleet usein sellaisia, joiden semanttinen preferenssi ei ole odotuksenvastainen tai edes yllättävä. Tällainen on jo luvussa 2.3.4 esiin tuotu Sinclairin (1996) *naked eye* -ilmausta koskeva analyysi: ei voitane pitää yllättävänä, että ilmauksen semanttiseksi preferenssiksi osoittautui 'näkyminen, näkyvyys'; vrt. suomen kielen *paljain silmin*. Samankaltaista ennalta-arvatta-vuutta on nähtävissä *true feelings* ('todelliset tunteet') -tapauksessa (mts. 89–90), jonka preferenssiksi osoittautuu 'näyttäminen, osoittaminen'.

Sen sijaan merkitykseltään abstraktien ilmausten semanttisen preferenssin analyysi ei tuone esiin yhtä selkeitä semanttisia valintoja, vaan kontekstuaalinen valinta osoittautuu nimenomaan preferenssiksi, ei aukottomaksi valintarajoitukseksi. Tällainen on tilanne ainakin merkitykseltään epätäsmällisillä astemääritteillä. Tämä selviää jo luomalla silmäys kollokaattianalyysin yhteydessä olevaan taulukkoon 6.5 (s. 91). Esimerkiksi *hyvin*-määritteellä on pelkästään taajimmin esiintyvien kollokaattien joukossa sanoja, joiden merkitykset ovat mitä erilaisimpia ja toisinaan hyvin abstrakteja, mikä vaikeuttaa merkitysten luokittelua analyysin kannalta järkeviin luokkiin: esimerkiksi *olla* ('olemassaolon, omistamisen' jne. ilmaiseminen), *ja* ('yhdistäminen'), *minä* ('itseän viittaava'), *sanoa* ('ilmaista puhumalla'), *mies* ('aikuisen miessukupuolta oleva ihminen') ja *kaunis* ('ulkoisilta ominaisuuksiltaan viehättävä'). Näille ja kaikille *hyvin*-määritteen tarkastelualueen 1760 saneelle olisi mahdotonta löytää yhtä tai edes muutamaa merkitysluokkaa, johon lekseemit olisi mahdollista tyhjentävästi sijoittaa. Sama pätee myös muihin astemääritteisiin ja niiden kotekstiin. Sen vuoksi tarkastelussa on tyydytty edellisistä analyyseista poiketen analysoimaan vain O1-positiossa olevat saneet ja ryhmittelemään ne merkitysyryhmiin.

Semanttisen preferenssin analyysille on luonteenomaista, että valintapreferenssiä etsitään sanaluokkarajat ylittävästi (ks. tarkemmin lukua 2.3.4.1). Tämän vuoksi kotekstin lekseemeille olisi löydettävä sellaisia luokituksia, jotka mahdollistavat sanojen ryhmittelemisen merkityksittäin siten, että sanaluokkarajat ja niiden sisäiset luokittelut eivät ole preferenssianalyysin esteenä. Huolimatta siitä, että seuraavaksi tehtävä analyysi on jo rajattu O1-positioon, ovat luokiteltavien sanojen merkitykset monenkirjavia, sillä positiossa on niin kvanttoreita, adpositioita, adverbeja kuin adjektiivejakin, joista etenkin kahdessa jälkimmäisessä ryhmässä on merkityskirjo suuri. Kyseisessä asemassa on merkitykseltään selkein ryhmä kvanttorit, jotka ilmaisevat pääasiassa määrää ja järjestystä. Luokka on jaettavissa edelleen alaryhmiin universaalit (*jokainen, kaikki*), eksistentiaaliset (*joku, eräs*), epämääräistä määrää ilmoittavat (*muutama, harva*), vertailevat (*sama, muu*) ja kielteisiin yhteyksiin rajoittuvat (*kukaan, kumpikaan*) kvanttorit (Vilkuna 1996: 52). (Vilkunan jaottelun perusteet ovat nähtävissä jo Tuomikoskella 1969; samoilla linjoilla ovat osittain myös Hogg 1977 ja Frawley 1992.) Hakulinen ja Karlsson (1979: 81–82) ryhmittelevät kvanttorit syntaktisin perustein kuitenkin hieman toisin: lukusana- (*kaksi, monta*), lukumäärä- (*harva, toinen*), yksilö- (*eräs, ei kukaan*), paljous- (*hiukan, kaikki*) ja monikolliset (*eräät, useat*) kvanttorit. Kaikille alakategorioille on kuitenkin

kin yhteistä kvantifiointi, eikä kvanttoreiden sisäinen kategoriointi ole ratkaisevaa käsillä olevan työn kannalta – tosin tässä työssä käytettävät vahventavat astemääritteet intensifioivat tyypillisesti vain epämääräistä määrää ilmoittavia kvanttoreita (Tuomikosken kvantitatiivisia ja Hakulisen ja Karlssonin jaottelussa soveltuvin osin lukumäärä-, paljous- ja monikollisia kvanttoreita).

Adpositioina käytetään suomessa mm. paikkaan (*sillan alla*), suuntaan (*kohti kuolemaa*), aikaan (*viikon päästä*), tapaan (*ihmisten mukana*) tai karitiivisuuteen (*ilman rahaa*) liittyviä tai muunlaista suhdetta ilmoittavia (*tästä huolimatta, siihen nähden*) ilmauksia (Hakulinen–Karlsson 1979: 154–156; ks. adpositioista myös esim. Quirk ym. 1985). Tämän luokan kanssa yhteneviä merkitysryhmiä on adverbeilla, joista erityisesti materiaaliset adverbit (Ahlman 1933) voivat olla astemääritteiden pääsanoina. Ryhmään luetaan Ahlmanin (mts. 142–144) jaottelua noudattaen paikkaa (*kotona, lähellä*), aikaa (*eilen, pian*), tilaa (*ahtaalla, hengissä*) ja tapaa (*kauniisti, hyvin*) ilmoittavia ja näiden sukuisia seikkoja. (Samansuuntainen jaottelu on myös Jacobsonilla (1964: 21–25, 76–86), joskaan tämä ei noudata Ahlmanin tapaan yläkategorisointia materiaalsiin, modaalisiin ja intensiteettiadverbeihin.)

Edellisiä luokkia monimutkaisempaa lienee kuitenkin adjektiivien luokittelu, eikä vähiten luokkaan kuuluvien sanojen runsaslukuisuuden vuoksi⁶². Dixon (1977: 31) luokittelee adjektiivit seitsemään ryhmään: ulottuvuus (suomessa esim. *iso, pieni, pitkä, lyhyt, laaja, paksu*), fyysinen ominaisuus (*kova, pehmeä, painava, kuuma hapan*), väri (*musta, punainen*), inhimillinen taipumus (*onnellinen, viisas, julma, ylpeä*), ikä (*nuori, vanha*), arvo (*hyvä, paha*) ja vauhti (*nopea, hidas*). Jaottelu perustuu englannin tyypillisimpien adjektiivien analyysiin (mts. 30–31) ja sitä on käytetty typologisen kuvauksen pohjana. Luokittelua jatkaa edelleen Stassen (1997: 168) lisäämällä inhimillisen taipumuksen, fyysisen ominaisuuden, ulottuvuuden, iän, värin ja arvonn rinnalle muodon (*pyöreä, suora*), materiaalin (*puinen, hopeinen*) ja sukupuolen (*miehininen, naisellinen*) ja pudottamalla Dixonin vauhdin luokittelusta. Edelleen Dixoniin nojautuen adjektiiveja on luokitellut Frawley (1992), joka pitää kuitenkin inhimillistä taipumusta osoittavien adjektiivien ryhmää liian heterogeenisenä ja hajanaisena (ks. myös Wetzer 1996: 9). Hän ehdottaakin ryhmää jaettavaksi henkistä (*onnellinen, viisas, mustasukkainen*) ja fyysistä (*heikko, janoinen*) tilaa ja käyttäytymistä (*villi, laiska*) kuvaaviin adjektiiveihin (mts. 458). Myös Dixonin fyysinen ominaisuus saa tarkemman jaottelun: aistimus (*äänekäs, suolainen*), koostumus (*pehmeä, kuminen*), pintarakenne (*karkea, pehmeä*), lämpötila (*kylmä, lämmin*), syöminen (*raaka, keitetty*), hahmo (*terävä, rikkinäinen*) ja tiiviys (*painava, kokonainen*).

Yhden adjektiivien ryhmittelymallin tarjoavat vielä Leppäjärvi ja Jämsä (1976: 5–10) suomen 1960-luvun adjektiiveja koskevassa tutkimuksessaan: heidän luokittelunsa perustuu adjektiivien pääsanana tarkoitetun mukaiseen jaotteluun, eli pääsanana merkitys (joko fyysisyys eli pääsanana tarkoite on ihmistajunnan ulkopuolella tai psyykkisyys eli pääsanana kuuluu ihmisen tajuntaan) toimii ryhmittelyn perusteena toisin kuin edellä. Tässä luokittelussa adjektiivit on kategorioitu reseptiivisiin (aistivaikutelma), lokaalisiin (paikka, asema), statuaalisiin (tila, koostumus), mensu-

⁶² Luokkaan luettavien sanojen määrä vahtelee luonnollisesti sen mukaan, onko kielessä kysyinen luokka ja onko luokka avoin vai suljettu. Suomessa luokka on avoin ja adjektiivien määrä runsas.

raalisiin (mitattavuus), temporaalisiin (aika), komparatiivisiin (vertailevat), possessiivisiin (tarkoitteen omistaminen) ja emotionaalisiin (asenne) adjektiivihin.

Kuten huomata saattaa, adjektiivien luokittelu on monenkirjavaa, osin päällekkäistä ja ongelmallista. Pajunen (2000: 47, ks. myös 53–) on kiinnittänyt huomiota luokittelun problematiikkaan ja huomauttaa esimerkiksi, että edellä mainituista Dixonin ryhmittelyyn perustuvista piirteistä muoto voidaan käsittää ulottuvuuteen tai fyysiseen ominaisuuteen liittyväksi ja että fyysisiä ominaisuuksia voidaan pitää myös inhimillisinä. Edelleen voitaisiin pohtia myös aistimus-piirteen suhdetta muihin tyypeihin: eivätkö muun muassa ulottuvuus, väri, muoto, materiaali, lämpötila ja monet muut ominaisuudet ole havaittavissa juuri aistimisen kautta? Jaottelut semanttisiin tyypeihin jäävät siis osittain yhteismitattomiksi; tyypejä voitaisiin pitää samanarvoisina tai niitä saatettaisiin ryhmitellä erilaisten yläkategorioiden alle hyvin monella tapaa.

Semanttisen preferenssin analysoimisen kannalta olisi ratkaistava vielä eri sanaluokkiin kuuluvien sanojen ryhmittely sanaluokkarajat ylittäviin semanttisiin ryhmiin. Edellä kuvattu ryhmittelyjen moninaisuus – etenkin adjektiivien osalta – tekee valmiiden semanttisten ryhmien luomisen jossain määrin ongelmalliseksi. Joitakin selviä ryhmiä voidaan kuitenkin luoda analyysin pohjaksi; taulukossa 6.24 on kuvattu tiivistetysti edellisiä ryhmittelyjä. Siitä on luettavissa joitakin eri sanaluokkakategorioita yhdistäviä semanttisia ominaisuuksia. Tällaisia ovat suhteellisen selvästi paikka (lokaalisuus) ja aika (temporaalisuus). Edelleen myös kvantiteetti on läsnä jossain määrin kaikissa luokissa: kvanttoreissa, adpositioiden karitiivisessa käytössä, intensiteettiadverbeissa ja adjektiiveista mahdollisesti esimerkiksi ulottuvuutta, ikää, lämpötilaa ja tiivyyttä merkitsevien luokissa. Toisaalta ulottuvuus liittyy läheisesti myös muotoon, ikä temporaalisuuteen ja lämpötila aistihavaintoihin. Jälleen siis luokittelu osoittautuu monimutkaiseksi.

Semanttisen preferenssin analyysissä on tavoitteena löytää suhteellisen hienojakoisia merkityspiirteitä (kuten edellä mainittu Sinclairin 'näkyminen, näkyvyys'), mutta toisaalta on tarkoituksenmukaista pyrkiä tarpeeksi pieneen määrään ryhmiä, jotta analyysi on hallittavissa. Luokittelu jää siten pakostakin kutakuinkin läheisten merkityssisältöjen yhteenniputtamiseksi. Oletukseni kuitenkin on, että näin aste-määritteiden semanttista preferenssiä pystytään kuvaamaan hallitummin kuin laventamalla erilaisten semanttisten ryhmien kirjo hyvin laajaksi.

Olen päätenyt luomaan edellisen kuvauksen perusteella analyysini pohjaksi seuraavat sanaluokkajaon ylittävät merkitysluokat: paikka, aika, määrä. Tämä ei luonnollisestikaan ole riittävä ryhmittely preferenssin kuvauksen pohjaksi. Etukäteen tehtävän ryhmittelyn problemaattisuuden vuoksi ole päätenyt osittain soveltamaan korpusvetoista lähestymistapaa. Tälle menetelmälle on tyypillistä nostaa aineisto ensisijaiseen asemaan, sen sijaan että aineistosta etsittäisiin tukea jo olemassa olevalle teorialle. Aineistoa ei siis käytetä pelkästään esimerkivarastona, josta etsitään aiempaan tutkimukseen soveltuvia esimerkkejä, vaan toistuvat esiintymät ja säännönmukaisuudet nostetaan kielellisen kuvauksen pohjaksi, jolloin aineisto nousee hypoteesien ja teoreettisen kuvauksen lähtökohdaksi. (Menetelmästä Tognini-Bonelli 1996: 69–72; ks. myös 2001 luvut 5 ja 6.) En kuitenkaan tee semanttisen preferenssin analyysiani täysin korpuslähtöisesti, vaan mikäli aineistosta on löydetävissä selvästi edellä mainittuja merkitystyyppisiä, käytän ryhmittelyssä apuna näitä

valmiita tyyppejä. Toisaalta en kuitenkaan anna etukäteen tiedossa olevien ryhmitelyjen kahlita analyysia, vaan ryhmittelen aineistoa nimenomaan sen mukaan, millaisiin jaotteluihin aineisto antaa aihetta.

Taulukko 6.24. O1-positiosta tavattavien luokkien semanttisia ryhmittelyjä aiemman kirjallisuuden perusteella.

adjektiivit	adverbit	adpositiot	kvanttorit
Dixon (D, 1977), Frawley (F, 1992), Stassen (S, 1997), Leppäjärvi–Jämsä (LJ, 1976)	Ahlman (1933)	Hakulinen–Karlsson (1979)	Vilkuna (1996)
Inhimillinen taipumus (D)	Materiaaliset	Paikka	Universaaliset
-Henkinen tila (F)	-Paikka	Suunta	Eksistentiaaliset
-Fyysinen tila (F)	-Aika	Aika	Epämääräistä määrää
-Käyttäytyminen (F)	-Tapa	Tapa	ilmoittavat
Fyysinen ominaisuus (D)	-Tila	Karitiivisuus	Vertailevat
-Aistimus (F)	Intensiteettiadv.	Muut suhteet	Kielteisiin yhteyksiin rajoittuvat
-Koostumus (F)	Modaaliset		
-Pintarakenne (F)			
-Lämpötila (F)			Kaikissa kvantifiointi
-Syöminen (F)			
-Hahmo (F)			
-Tiiviys (F)			
Ulottuvuus (D)			
Ikä (D)			
Väri (D)			
Arvo (D)			
Vauhti (D)			
Muoto (S)			
Materiaali (S)			
Sukupuoli (S)			
Lokaalisuus (LJ)			
Temporaalisuus (LJ)			
Komparatiiviset (LJ)			
Possessiiviset (LJ)			
Emotionaalisuus (LJ)			

Luokittelu on tehty kollokaattien tekstiympäristöstä syntyvän merkityksen perusteella, jolloin huomioon on otettu kollokaatin merkitys nimenomaan kulloisessakin kontekstissa. Niinpä esimerkiksi *herkkä* on luettu fyysistä aistimusta ilmentäväksi (konkordanssi 6.21 rivi 1), vaikka toisissa yhteyksissä se saattaisi kuvata myös henkistä tilaa tai luonteenpiirrettä, *läheinen* vuorostaan kuvastaa kontekstissaan (rivi 2) emotionaalisia piirteitä, vaikka toisaalla sitä voidaan käyttää myös paikkaan viittaavana (*läheinen puistikko*), saman tulkinnan saavat myös *soma* ja *tyyni* (rivit 3 ja 4). Rivin 5 *hauraus* on puolestaan tässä tapauksessa (metaforisen) kieliopin kelvollisuuteen tai hyvyyteen liittyvä, ja se on ryhmitelty arvoa merkitseväksi.

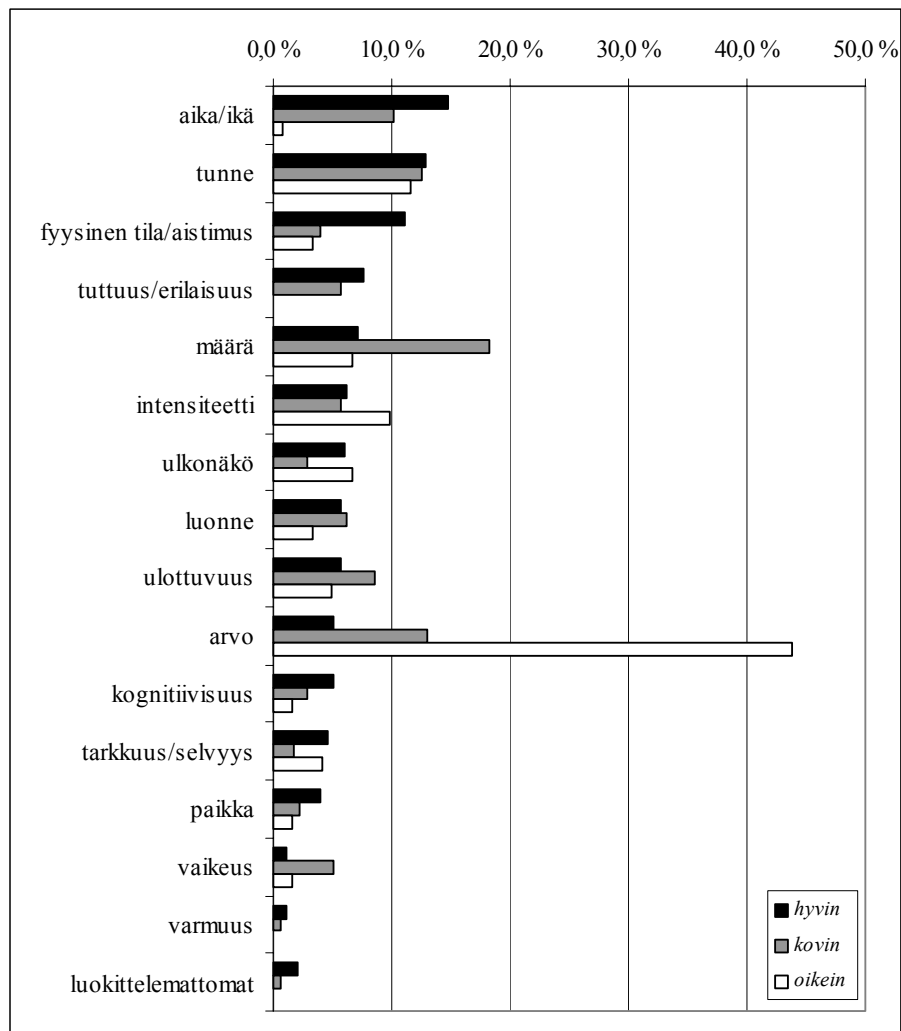
Samakin ilmaus voi toisinaan kontekstinsa perusteella tulla luokitelluksi eri ryhmiin: *kovasti* on laskettu määrää ilmaisevaksi rivillä 6 mutta intensiivistä tekemisen tapaa ilmentäväksi rivillä 7. Kukin luokka sisältää myös antonyymiparit. Siten esimerkiksi arvoa ilmentävät sisältävät muun muassa ilmaukset *hyvä* ja *hyvin* mutta myös sanat *huono*, *huonosti* ja *paha* (rivit 8–13).

Aina kollokaattien ryhmittely ei ole ollut ongelmatonta, sillä toisinaan sanan voisi sijoittaa useampaankin luokkaan kuuluvaksi. Tällainen ongelma on muun muassa *puhtaus-* ja *viehättävä-*kollokaattien kanssa: Puhtaus ei ole välttämättä pelkästään näköaistin avulla havaittava asia, vaan sen voi aistia myös haistamalla tai tuntemalla, jolloin olisi perusteltua luokitella fyysistä aistimusta ilmentävään luokkaan. Rivillä 14 oleva tapaus kuvastaa kuitenkin selvästi näköaistin perusteella tehtyä havaintoa, joten tapaus on luokiteltu ulkonäköä merkitsevään ryhmään. Rivin 15 *viehättävä* on myös ongelmallinen sijoitettava, liittyyhän siihen muun muassa miellyttävyyden merkityspiirre, eli se on vahvasti myös emotionaalinen sana. Kyseisessäkin tapauksessa viehättävyys tulee kuitenkin aistituksi nimenomaan näkemisen kautta, joten sana on lueteltu ulkonäköä merkitsevien sanojen joukkoon. Rivin 16 *viehättävä* kuvaa taas luonteenpiirrettä, joten se on ryhmitelty luokkaan luonne, ja rivillä 17 on kyse tunteesta.

Konkordanssi 6.21. Esimerkkejä lekseemien ryhmittelystä merkitysluokkiin.

1	sta velvollisuuksista. Hänen ihonsa oli	<i>hyvin</i>	herkkä . Hän hämmentyi, kun taputtelin
2	n neiti Rosen ollut minulle vuosikautia	<i>hyvin</i>	läheinen ihminen - äidin ohella joka oli
3	skeleen. Kepeästi. Elämä on hetkittäin	<i>hyvin</i>	somaa . Puukengät kopisivat hyvinkin
4	soa niin tuttua. Ja Asko oli sisäisesti	<i>hyvin</i>	tyyni , vaikka hänen liikutuksensa oli
5	hen uskoin yhä, mutta sen kielioppi oli	<i>kovin</i>	hauras katseitten ja korvien alla. He v
6	lta, hän keksi vaihtoa puheenaihetta. -	<i>Oikein</i>	kovasti terveisiä, hän haluaisi tavata
7	vä liikemieskin voi olla rehellinen, jos	<i>oikein</i>	kovasti yrittää, hänen isällään oli ollut
8	. Marjatta sanoi aina että Elinasta tulisi	<i>oikein</i>	hyvä yksikätinen pianisti, mutta kun v
9	kilvan ja äänekkäästi, että ymmärsivät	<i>oikein</i>	hyvin liikuttumisen eikä siinä ollut mi
10	onosti. Ja vielä huonommin. Ja oikein	<i>hyvin</i>	huonosti . Minä voin istua tunkiolla ra
11	eläimiin, mutta apinan tähtääminen on	<i>kovin</i>	huonoa . Tämä asia ole palkinnon mot
12	suomalaiset ystävänsä olleet joutuneet	<i>kovin</i>	pahaan tilanteeseen autettuaan häntä l
13	net pelastusveneeseen ja sanoivat: - Jos	<i>oikein</i>	paha paikka tulee, Caleb voi rukoilla
14	yli sata säntillisissä riveissä. Salissa oli	<i>hyvin</i>	puhdasta , melkeinpä pelottavan kiiltäv
15	e näkisi hyvin. Sarin mielestä aitta oli	<i>oikein</i>	viehättävä paikka kynttilänvalossa. Sa
16	in yleisesti ottaen olen pohjimmiltani	<i>hyvin</i>	viehättävä mies, ja naiset rakastuvat m
17	inua onnenkypäläkseen, minusta se on	<i>hyvin</i>	viehättävää . Minulta hän sai kirjansa il

Korpusvetoisen ja aiempia luokitteluja soveltavan ryhmittelyn jälkeen O1-kollokaatit muodostivat aineistossa seuraavat luokat: aika (jossa myös ikä), arvo, fyysinen tila/aistimus, intensiteetti, kognitiivisuus, luonne, määrä, paikka, tarkuus/selvyys, tunne, tuttuus/erilaisuus, ulkonäkö, ulottuvuus, vaikeus ja varmuus. Lisäksi analyysiin jouduttiin lisäämään luokittelemattomien luokka, koska ryhmittelystä oli mahdotonta tehdä täysin aukotonta. Analyysin perustana oleva O1-kollokaattien semanttinen luokittelu eri ryhmiin on esitetty liitteessä 2. Astemääritteiden semanttiset preferenssit näyttävät positiossa O1 olevien kollokaattien merkitysten perusteella kuviossa 6.6 esitetyn kaltaisilta.



Kuvio 6.6. Semanttinen preferenssi positiossa O1.

Merkityspiirteistä aika/ikä ja tunne ovat määritteillä suhteellisen yleisiä, paikka, vaikeus ja varmuus vuorostaan harvinaisia – näissäkin ryhmissä on kuitenkin yleisyseroja astemääritteiden välillä. Käsitteiden astemääritteiden preferenssejä seuraavaksi määritteittäin.

hyvin-määritteellä on muita voimakkaampi preferenssi fyysisistä tilaa/aistimusta ilmaisevaan merkityspiirteeseen (11,1 %, n = 39) ja toisaalta selvä negatiivinen

preferenssi arvoon (5,1 %, n = 18)⁶³ suhteessa muihin määritteisiin. Jälkimmäinen johtuu pitkälti siltä, että kollokaatiot *hyvin* <HYVÄ, *hyvin*> ovat toisteisuutensa vuoksi erittäin harvinaisia, mutta näyttää siltä, että *hyvin* ei määritä myöskään muita arvo-piirteeseen luettuja sanoja. Alla on lueteltu esimerkkejä fyysistä tilaa ja aistimusta merkitsevistä kollokaateista:

hyvin <KIIHOTTAVA, KUOLLUT, KYLMÄ, LÄMMIN, VÄSYNYT>.

kovin-määritteellä on synonyymejaan merkitsevästi suurempi preferenssi piirteeseen määrä (18,2 %, n = 32).⁶⁴ Muilla tämän piirteen osuus jää selvästi alle kymmenen prosentin. Esimerkkejä tätä piirrettä ilmaisevista kollokaateista:

kovin <MONI, niukasti, paljon, vähän>.

oikein-määritteellä on muista poiketen yksi erittäin voimakas semanttinen preferenssi: arvo kattaa koko merkityskirjosta peräti 43,8 % (n = 53)⁶⁵. Preferenssi muodostuu lähes yksinomaan kollokaateista HYVÄ ja *hyvin*, jotka myös olivat tämän määritteen kollokationaalissa valinnassa selvästi muita yleisempiä. Muita tähän merkitysluokkaan sijoitettuja kollokaatteja ovat

oikein <KURJA, MAINIO, ONNISTUNUT, PAHA>.

Selvät negatiiviset preferenssit *oikein*-määritteellä on synonyymeihinsa verrattuna vuorostaan merkityspiirteisiin aika/ikä ja tuttuus/erilaisuus⁶⁶. Näiden lisäksi astemääritteparin *hyvin* ja *kovin* jäsenet eroavat toisistaan vielä preferenssiltään arvo- ja vaikeus-piirteisiin: *kovin*-määritteellä on näihin voimakkaampi preferenssi⁶⁷. Vaikeus-piirteen ilmentymiä on alla:

kovin <HANKALA, HELPPO, helposti, VAIKEA, YKSINKERTAINEN>.

Vaikka tässä luokitellut semanttiset preferenssit jakautuvat kokonaisuutena melko tasaisesti synonyymien välillä, erottuu synonyymikohtaisia selviä preferenssejä siis muutama. Tällaisiksi oli osoitettavissa *hyvin*-määritteen semanttinen preferenssi fyysinen tila/aistimus, *kovin*-määritteen määrä (ja vaikeus) sekä *oikein*-määritteen arvo. Tämän lisäksi synonyymeja erottavat muutamat negatiiviset preferenssit.

⁶³ fyysinen tila ja astimus: *hyvin* vs. *kovin* z = 2,727970 (p < 0,01); *hyvin* vs. *oikein* z = 2,566004 (p < 0,05), arvo: *hyvin* vs. *kovin* z = 3,219612 (p < 0,01); *hyvin* vs. *oikein* z = 10,27848 (p < 0,001).

⁶⁴ *hyvin* vs. *kovin* z = 3,867396 (p < 0,001); *oikein* vs. *kovin* z = 2,869944 (p < 0,01).

⁶⁵ *hyvin* vs. *oikein* z = 10,27848 (p < 0,001); *kovin* vs. *oikein* z = 5,963976 (p < 0,001).

⁶⁶ aika/ikä: *hyvin* vs. *oikein* z = 4,195565 (p < 0,001); *kovin* vs. *oikein* z = 3,253075 (p < 0,01), tuttuus/erilaisuus: *hyvin* vs. *oikein* z = 3,137375 (p < 0,01); *kovin* vs. *oikein* z = 2,667311 (p < 0,01).

⁶⁷ arvo: *hyvin* vs. *kovin* z = 3,219612 (p < 0,01), vaikeus: *hyvin* vs. *kovin* z = 2,780063 (p < 0,01).

Määritteille on siis löydettävissä niitä toisistaan erottavia semanttisia preferenssejä, vaikka tarkastelussa keskityttiin vain yhteen positioon eli astemääritteiden pääsanoina oleviin kollokaatteihin. Tässä analyysissä luokkia muodostui aineiston perusteella 15 (ja lisäksi luokittelemattomien ryhmä), mutta luokitusta olisi mahdollista muuttaa myös hienojakoisemmiksi, jolloin semanttisesta preferenssistä muodostuisi tarkempi kuva. Analyysistä näkyy myös, että näin abstrakteille leksikaalisille yksiköille ei löydy yhtä, tai edes kahta selvää semanttista preferenssiä, kuten Sinclairin esimerkeissä. Astemääritteistä frekventeimmät lienevätkin eräänlaisia yleismääritteitä, jotka eivät ole selvästi sidoksissa johonkin tiettyyn preferenssiin vaan pystyvät määrittämään hyvin erilaisia pääsanoja. Harvinaisemmilla astemääritteillä tilanne saattaa olla jo toisenlainen; tästä antaakin viitteitä jo *oikein*-määrite, jolla on muita selvempi preferenssi yhteen piirteeseen ja joka on tässä käsitellyistä määritteistä aineistossa harvinaisin. Määritteiden preferenssikirjoon vaikuttaa myös astemääritteiden kieliopillinen luonne: onhan oletettavaa, että frekventeillä kieliopillisilla sanoilla ei ole kovin selkeitä semanttisia preferenssejä.

6.4.2. Semanttinen prosodia: positiivinen ja negatiivinen aura

Seuraavaksi analysoidaan määritteiden prosodiset ominaisuudet keskittyen kontekstin positiivisiin ja negatiivisiin merkityspiirteisiin. Olen ottanut tarkasteluun tarvittaessa alkuperäistä havaintoaluetta V3–O2 suuremman alueen, koska toisinaan lähimmästä kontekstista ja jopa lause- ja virketasonkin analyysistä voi olla mahdoton päätellä, onko leksikaalisen yksikön prosodia negatiivinen vai positiivinen. Tällöin aluetta joudutaan laajentamaan tarkastelualuetta ja virkettäkin suuremmaksi.

Vaikka määrittämissuhde saattaakin lähikontekstin perusteella tulla ensin tulkituksi positiivista prosodiaa osoittavaksi, voi kontekstin laajentaminen paljastaa prosodian päinvastaiseksi. Seuraavassa esimerkissä tavattava *puhtaus* ja erityisesti se, että on hyvin puhdasta, tulkittaneen yleisesti positiiviseksi. Kuitenkin esimerkkitapauksessa *puhtaus* koetaan ahdistavana ja liiallisena, mikä paljastuu tarkastelualueen (hakasulkeissa) jälkeisestä oikeasta kontekstista. Tällaiset tapaukset korostavat tarkastelualueen laajentamisen merkitystä leksikaalisten yksiköiden prosodian tutkimisessa.

Louner veti salin ilmaa syvälle sieraimiinsa. Huone tuoksui aseptisen puhtaalta, kliiniseltä ja hyvin valkoiselta. Vuoteita oli varmasti yli sata säntillisissä [*riveissä*]. *Salissa oli hyvin puhdasta, melkeinpä*] pelottavan kiiltävää. Mutta eniten Louner oudoksui tätä pysähtynyttä hiljaisuutta. Ei puhetta, ei inahdustakaan missään.

Pelkän tarkastelualueen analysointi voikin johtaa pahasti harhaan. Tästä käy myös esimerkiksi hakasulkeilla merkitty tarkastelualue seuraavassa esimerkissä, jossa *hyvin* määrittää *elävä*-adjektiivia. Nukkien elävältä näyttäminen voi tulla tulkituksi negatiiviseksi, koska seuraavan lauseen sana, joka yltää tarkastelualueelle,

on *yksinäinen*. Laajempi, tässä jopa useampivirkkeinen konteksti kuitenkin paljastaa, että nukkiin – tässä tapauksessa mallinukkiin – elävältä näyttäminen koetaan positiiviseksi ja konteksti heijastaa siten positiivista prosodiaa:

Katriina oli rakentanut työhuoneestaan verrattoman jatkopaikan. Portaikko oli koristeltu kankain alhaalta aina neljänteen kerrokseen. Portaikkoon ja työhuoneeseen oli rakennettu mallinukeista erilaisia [*ryhmiä. Nuket vaikuttivat hyvin eläviltä. Yksinäinen*] jatkovieras saattoi asettautua johonkin nukkeriryhmään ja sulautua siihen, niin ettei kukaan osannut häiritä.

Toisinaan sama sana tai samasta kannasta johdetut sanat voivat saada kontekstin perusteella erilaiset tulkinnat, mikä osaltaan tekee analyysistä haastavan. Seuraavien esimerkkien adjektiivin *uteliaas* ja adverbien *uteliaasti* olen tulkinnut esimerkissä a) positiiviseksi, esimerkissä b) negatiiviseksi, kun taas tapauksessa c) ei liene selvästi kyse kummastakaan tapauksesta.

- a) Hautausmaan keskellä oli kaunis goottilaisista tiilistä muurattu kirkko ja kirkon seinässä tiiliseinään kaiverrettu teksti, jonka oppaana toiminut katolinen pappi luki, ja käännettynä teksti, jos muistan oikein, kuului: "Puolalainen, muista että isänmaasi on Puola, mutta äitisi on Kasubia". Goottilainen tiili on kunnan tiili. Se on iso ja siitä kun muuraa niin kestää. Se sulhanen istui minua vastapäätä [*ja katsoi minua hyvin uteliaasti. Sarin*] poika alkoi räplätä televisiota auki. [--] Ne saattavat olla joskus kiusallista väkeä, sulhanen sanoi. Sieluni tuntui lennokkaalta ja ilmavalta. Monia ikimuistettavia asioita virtasi kuin puroina mieleeni --.
- b) Rouva Korhonen sanoi että voi ollakin. Minä sanoin että nykyisin näihin perinteisiin suhtaudutaan aika joustavasti. Minun oli vaikea puhua kun suuta kuivasi. Sanoisin rouva [*Korhosen katsoen minua hyvin uteliaana, ja*] sitten tunsin hikipisaroitten valuvan otsallani. Rouva Korhonen huomasi sen täysin varmasti. Näin sen hänen ilmeestään. Yritin livahtaa tilanteesta ja ajatella selkeästi.
- c) Minä kuulin toissa-aamuna radion aamuhartauden. Siinä puhuttiin aviorikoksesta, ja jäin miettimään sitä sanaa, sitä että se oikeudellisesti on rikos ja avioeroperuste, tai oli ennen. Oletko sinä tehnyt [*aviorikosta? Hän näytti hyvin uteliaalta. - En,*] minä sanoin. - Etkö? Ihan totta? Sanoin minua oikeastaan harmittavan etten ollut. Olin lukenut aikaisemmin Osmo Kontulan ja Elina Haavio-Mannilan tekemästä tutkimuksesta että --.

Tapaukset, joissa prosodian päättely ei ole mahdollista (c) tai jotka ovat mielestäni neutraaleja, on luokiteltu omaksi ryhmäkseen. Epäselvästä tapauksesta käy esimerkiksi myös seuraava:

Ne [sudet] ovat ovelia ja röyhkeitä, kun taas 1970-luvun ihmiset eivät ole enää yhtä valppaita kuin viime vuosisadan salojen asukkaat. [*Talveksi sudet hajaantuvat hyvin pieniksi parviksi*] ympäri Etelä-Suomea. Laukon peurat, hirvet, koirat ja rusakot ovat niiden ravintoa. Kevättalvella liittyy niihin kolme idästä Luumäen, Anttolan ja Joutsan kautta tullutta urossutta.

Esimerkin tarkastelualueesta kuten myöskään sitä ympäröivästä kontekstista ei pysty yksiselitteisesti päättämään, ovatko *hyvin pienet parvet* [sic!] negatiivisia vai positiivisia. Tulkintoja olisi nimittäin ainakin kaksi: susien kannalta pienet laumat ovat positiivinen asia, koska niiden liikkuminen, saalistus ja piiloutuminen on silloin helpompaa. Ihmisen ja saaliseläimen kannalta tilanne on kuitenkin päinvastainen. Tämänkaltaisissa tapauksissa joudutaan turvautumaan kielenulkoisen tilannekontekstin analysoimiseen, koska kontekstin muista leksikaalisista yksiköistä ei sinänsä tule suoraa tukea prosodian valintaan. Niinpä ratkaisu on tehtävä sen perusteella, mistä tekstissä puhutaan, miten asiaan, olioon tai ihmiseen suhtaudutaan. Joskus prosodisten piirteiden hahmottaminen saattaa siis vaatia jopa yleisen mielipiteen, maailmantiedon tai muun vastaavan huomioon ottamista. Tällöin kuitenkin liikutaan pitkälti jo tutkijan intuition varassa, ja toinen tekstin analysoija voisi päätyä toisenkinlaiseen tulokseen. Semanttisen prosodian, kuten edellä jo semanttisen preferenssinkin, tutkiminen tuo aineiston analyysiin muita kuvausmenetelmiä suuremman tulkinnan mahdollisuuden ja tutkijan oman intuition käytön.

Seuraavassa erittelyssä astemääritteiden prosodiaa on kuvattu käsitteillä *positiivinen*, *negatiivinen* ja *neutraali* (joka sisältää myös epäselvät tapaukset). Näistä negatiivinen prosodia on vielä jaettu kahtia: analyysissa on eritelty 1) tapaukset, joissa on selkeä leksikaalinen yksikkö, joka tuo negatiivisen prosodian esiin jollakin negatiivisena pidettävällä merkityssisällöllä, ja 2) tapaukset, joiden prosodia hahmottuu negatiiviseksi tilannekontekstin perusteella. Tällä jaottelulla olen tavoitellut tarkempaa kuvaa prosodisista ominaisuuksista. Negatiivisen merkityssisällön sisältäviksi olen lukenut muun muassa sellaiset tapaukset, jotka sisältävät

- a) kieltoverbin
- b) sanoja, joiden merkitys on negatiivinen (esim. adjektiivit, joissa on sanakielto, kuten EPÄMIELLYTTÄVÄ, EPÄTODELLINEN, tai jotka ovat merkitykseltään kielteisiä kuten adjektiivit VAIKEA, TYYTYMÄTÖN; adverbit kuten HUONOSTI, JULMASTI, VALITETTAVASTI; verbit ja verbirakenteet kuten KYLLÄSTYÄ, TAPPAA, OLLA OLEVINAAN; kieltohakuiset sanat kuten *lainkaan*, *mikään*)
- c) sanoja, jotka tekevät ilmauksen merkityksestä kielteisävyisen. Tällaisiksi olen lukenut sanat, jotka esiintyvät myönteisissä lauseissa, mutta muuttavat lauseen merkityksen (ainakin osittain) kielteiseksi, kuten adverbit *tuskin* ja *muka*.

Semanttisen prosodian analyysini noudattelee pääpiirteissään aiemmin Sinclairin (1998) kuvaamaa analyysia, jonka esimerkkinä oli englannin BUDGE-verbi ('hievahtaa'). Sinclair havaitsi verbin negatiivisen prosodian tulevan ilmi yhtäältä selvistä kieliopillisista negatioelementeistä (*not, can't, couldn't, won't, neither-nor*), mutta myös merkitykseltään negatiivisista verbeistä kuten REFUSE TO ('kieltäytyä'). Ellei selvää negatiivista elementtiä ollut havaittavissa negatiivisen prosodian perustaksi, oli prosodia kuitenkin hahmotettavissa BUDGE-verbin tilannekontekstista.

Yllä olevassa jaottelussani ne kotekstit, jotka sisältävät kieltoverbin, ovat kieltomuotoisia (lauseita), mutta kahdessa jälkimmäisessä tapauksessa koteksti voi olla sekä myöntö- että kieltomuotoinen. Sellaiset tapaukset, joissa kieltomuotoisessa koteksissa on kieltohakuinen kielenaines, kuten esimerkiksi *kukaan, mikään* (*Kukaan ei puhunut kovin lujaa*), ovat negatiivisen prosodian ilmitulon kannalta redundanteja, koska prosodia voidaan päätellä useasta sanasta. Tapauksia, jotka muuttuisivat myöntölauseen veroisiksi eli sisällöltään myönteisiksi eksplisiittisen kaksinkertaisen negaation (Hakanen 1973: 148–149) eli sekä lause- että sananegaation vuoksi (*Mies ei menetellyt epäviisaasti*)⁶⁸, on aineistossa vain yksi, konkordanssissa 6.22 tapaus rivillä 1. Tässä tapauksessa ilmauksen merkitys on myönteinen, koska laajemmasta kotekstista selviää, että henkilö nauraa onnellisuuttaan. (Kaksinkertaisesta kiellosta Hakanen mp.; Hakulinen ja Karlsson 1979: 271; myöntö- ja kieltohakuisista elementeistä ks. myös Sadeniemi 1942; Penttilä 1963; Kiuru 1977.)

Seuraavassa käsittelyssä nimitän edellä mainittuja kohtien a–c kielenaineiksi yhteisesti *negatioelementeiksi*. Käsite eroaa Penttilän (1963: 659–662) kieltolauseen tai kieltoyhteyden kuuluvista ilmausaineksista ja Savijärven (1984: 145–155) fakultatiivista kieltomorfeemeista siinä, että näillä on tarkoitettu nimenomaan kieltohakuisia liitteitä ja sanoja, jotka esiintyvät pääasiassa vain kieltomuotoisissa lauseissa ja joiden tehtävänä on muun muassa kiellon vahvistaminen. Negaatioelementti on hyvin väljä käsite, joka sisältää niin kieltoverbin, kieltohakuiset elementit kuin kotekstissaan kielteisen merkityksen sisältävät saneet ja ilmaukset, joilla ei ole kieltohakuisuusominaisuutta. Käsite on siis motivoitunut ilmausten ja koteksin merkityssisällön kautta, ei esimerkiksi muotovalintojen mukaan. Esimerkkejä negatioelementeiksi lasketuista kielenaineksista konkordanssissa 6.22.

⁶⁸ Hakulinen ja Karlsson (1979: 271) huomauttavat, että myöntömuotoinen lause (*Mies menetteli viisaasti*) ei ole tarkkaan ottaen synonyminen kaksinkertaisen negaation sisältämälle lauseelle. Tämä johtuu kaksinkertaisen kiellon varauksellisesta sävystä.

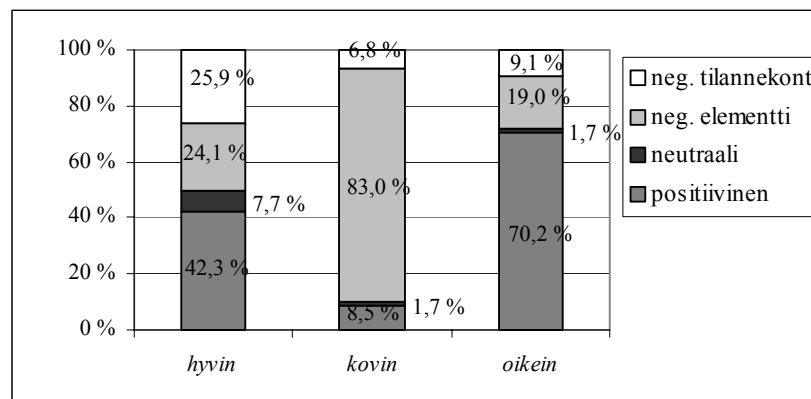
Konkordanssi 6.22. Esimerkkejä negaatioelementeistä.

1	n minä olin. Sitten hän nauroi: - En ole	<i>kovin</i>	onneton vieläkään. Sellainen hän ol
2	äsenten ydinjoukko nyt kuitenkin ei ole	<i>kovin</i>	ahnetta naurun päälle... Pienen met
3	, etteivät hänen suomalaiset ystävänsä olleet joutuneet	<i>kovin</i>	pahaan tilanteeseen autettuaan hänt
4	sa. Otin niistä toisen. Haistelín, ettei se	<i>kovin</i>	pahale löyhkäis. Olipa hyvä, että sa
5	tietenkään näkynyt. Arttu ei vaikuttanut	<i>kovin</i>	pettyneeltä , ja minä sanoin että min
6	mutettu. Ensimmäinen kohde odottaisi.	<i>Kovin</i>	röyhkeäksi hän ei varmaankaan uskaltaisi
7	ta saa. Hän ei ole kyllä koskaan pitänyt	<i>kovin</i>	surullisena näkynä lehtien putoami
8	ja orkussa mylvit kuin sonni. Ei se niin	<i>kovin</i>	vastenmieliseltä ole susia tuntunut.
9	en tulli ja rajamiehet eivät olleet meistä	<i>kovin</i>	kiinnostuneita, eivätkä venäläisetkä
10	in sillä tavalla, että tämä hallitus ei taida	<i>kovin</i>	kauan enää paikallaan istuskella. -
11	pintoja ajatellen tällainen vierailu ei olisi	<i>kovin</i>	hyvä juttu. - Älä kuule, Hujanen san
12	aivaalle oli kotkan nyt noustava. Mitään	<i>kovin</i>	liikuttavia jäähyväisiä ei jätetty. Kor
13	elimeen: ”Voikö tulla? Kullervo ei voi	<i>oikein</i>	hyvin.” Meillä jatkui Antin kanssa s
14	ada asiansa sanottua, mutta kun se tuntui	<i>kovin</i>	hankalalta , perääntyi sittenkin. Mik
15	nessa abstraktiudessaan herättää minussa	<i>hyvin</i>	epämiellyttäviä ja vihamielisiä ajatuksia.
16	irkonkylän koulussa, niin minulle sattui	<i>oikein</i>	ikävä ja hankala oppilas, eräs Jari
17	myöskin hänen suhtautuvan minuun heti	<i>hyvin</i>	niikeästi . Ihmettelin oliko Sari eht
18	en. - Ai jai, taas pääsi rouva pistämään	<i>oikein</i>	pahasti . Mutta totellaan, totellaan t
19	än hän kertoi minulle vasten kasvoja ja	<i>hyvin</i>	inhottavaan, alentuvaan sävyyn.
20	n vai mikä se oli, kyllästyisi vierailuihini	<i>hyvin</i>	nopeasti, ja silläkin olisi sanansa s
21	livat omassa asioissaan, jauhoivat muka	<i>hyvinkin</i>	tähdellisiä yhdentekevyyksiä . Nuor
22	siinä, kaksi kalaa kuivalla maalla, muka	<i>kovinkin</i>	erilaisia, toinen väittää olevansa tap

Aineiston analyysin jälkeen astemääritteiden prosodiat näyttävät kuviossa 6.7 esitetyn kaltaisilta. Tulos kertoo, että tarkasteltujen kolmen astemääritteen prosodiat ovat hyvin erilaisia. Ensinnäkin *hyvin*-määritteellä ei ole selvää positiivista tai negatiivista prosodiaa, sillä sen konteksti on lähes yhtä usein positiivista (42,3 %) ja negatiivista (50,0 %) prosodiaa ilmentävä. *kovin* poikkeaa synonyymistaan selvästi: sillä on erittäin voimakas negatiivinen prosodia, sillä noin 90 prosenttia tapauksista on negatiivista prosodiaa osoittavia. Näistä negaatioelementeillä – tyypillisesti kieltoverbilla – muodostettuja on ylivoimaisesti suurin osa. *oikein* on taas lähes *kovin*-määritteen vastakohta: negatiivisen prosodian osuus on vain noin 28 %, josta noin kaksi kolmasosaa on negaatioelementillä muodostettuja. Vaikka *oikein*-määritteellä on selvä positiivinen prosodia, se ei ole kuitenkaan niin vahva, kuin olisi voinut olettaa – löytyihän tämän määritteen kontekstista runsaasti HYVÄ- ja *hyvin*-kollokaatteja, ja sen semanttisen preferenssinkin analyysi nosti nimenomaan (positiivisen) arvon tyypilliseksi semanttiseksi preferenssiksi. Huolimatta positiivisesta prosodiasta määrite esiintyy siis jossain määrin myös negatiivissävytteisessä kontekstissa.

Näistä kolmesta astemääritteestä *hyvin* näyttää olevan eräänlainen ”yleismäärite” semanttisten ominaisuuksien perusteella, sillä sillä ei ole omaa synonyymeistaan erottuvaa semanttista preferenssiä eikä sen semanttinen prosodia osoita valintaa suuntaan eikä toiseen.

Edellä olevassa analyysissä negatiivista prosodiaa ilmentävien piirteiden jakaminen joko negatiivisesta elementistä tai tilannekontekstista johtuviksi osoittautui kannattavaksi. Varsinkaan *hyvin*-määritteen kontekstista ei läheskään aina löytynyt selvää negatiivista elementtiä, jonka perusteella kielteiseltä tuntuva tapaus olisi voitu laskea prosodialtaan kielteiseksi. Sellainen löytyi kuitenkin hyvin usein laajemman tilannekontekstin kautta, tapausten määrä oli peräti n. 26 % koko *hyvin*-määritteen esiintymistä ja noin puolet niistä tapauksista, joissa prosodia oli tulkittavissa negatiiviseksi. *kovin*-määritteellä tilannekontekstin avulla tehdyn päätelmän osuus on selvästi pienempi negatiivista prosodiaa osoittavista tapauksista; *oikein* sijoittuu näiden välille. Tämän tutkimuksen valossa näyttää siltä, että tarkastelualueetta on laajennettava tarpeen mukaan ohi ennalta määrätyn alueen, jolloin kuvauksessa voidaan hyödyntää laajempi konteksti. Tarvittaessa on myös turvaututtava tarvittaessa tekstinulkoiseen kontekstiin, jopa kielenkäyttäjien yleisiin mielipiteisiin, joiden avulla voidaan ryhtyä selvittämään prosodian olemusta.



Kuvio 6.7. Synonyymien prosodiat.

6.5. Kokoava katsaus: määritteiden *hyvin, kovin ja oikein* kontekstuaaliset profiilit ja synonyymisyys

Aiempaan tutkimuskirjallisuuteen perustuen työssä oletettiin, että synonyymiset ilmaukset ovat erotettavissa toisistaan ensiksikin kollokationaalisten (mm. Halliday 1966; Leech 1981; Persson 1989, 1990; Jantunen 2001b) ja kolligationaalisten (Biber ym. 1988; Jantunen 2001b) piirteiden avulla. Näistä tutkimusta laajennettiin edelleen kontekstin semanttisten ja morfologisten ominaisuuksien analysointiin. Niin semanttinen preferenssi (Sinclair 1996, 1998) kuin semanttinen prosodiakin (Louw 1993; Sinclair 1991, 1996; ks. myös Stubbs 2001a) on todettu leksikaalisen yksikön kontekstuaaliseen valintaan kuuluviksi piirteiksi, ja tässä työssä synonyymien oletettiin eroavan myös näiltä piirteiltään toisistaan. Tämän lisäksi esitin, että synonyymit voivat poiketa toisistaan myös morfologisen preferenssin suhteen, jonka postuloin sekä omien (Jantunen 2001b) että Arppen (2001, 2002) havaintojen pohjalta yhdeksi osaksi leksikaalisen yksikön kontekstuaalista valintaa. Aiemman tutkimukseni pohjalta ehdotin morfologisen preferenssin todellistuvan sekä paradigmaattisena eli lekseeminsisäisenä että kontekstuaalisena eli kontekstin leksikaalisissa yksiköissä esiin tulevana ominaisuutena.

Edellä kuvatussa analyysissä on jokaisessa muuttuja-analyysissä löytynyt piirteitä, jotka sekä *yhdistävät* astemääritteitä eli osoittavat ne synonyymisiksi että *erottavat* niitä toisistaan eli korostavat niiden yksilöllistä kontekstuaalista olemusta ja siten *lähisynonyymisyyttä*. Tarkastelussa on huomio kiinnittynyt pääasiassa synonyymeja erottaviin piirteisiin. On kuitenkin syytä ottaa käsiteltäväksi vielä yhdistävät eli synonyymisyyden muodostavat piirteet. Käsiteltyjen astemääritteiden synonyymisyys syntyy samanlaisen funktion ja käsitteellisen sisällön lisäksi useista yhdistävistä kontekstuaalisista ominaisuuksista. Synonyymisille ilmauksille – ainakin tässä käsitellyille – on mahdollista luoda niitä yhdistävä kontekstuaaliset preferenssit kuvaava profiili, joka esittää synonyymisyyttä korostavat piirteet kullakin kontekstuaalisella tasolla. Tässä tutkimuksessa samanlaisuutta kuvaava profiili muodostuu paljolti kontekstuaalisten piirteiden samanlaisesta kirjosta, kuten nähdään alla olevasta kuvauksesta.

Astemääritteille yhteiset piirteet eli niiden synonymiaa korostavat kontekstuaaliset valinnat:

- Kollokationaalinen valinta: yhteiset frekventit kollokaatit (kuten verbi OLLA, pronominit MINÄ, HÄN, SE, konjunktiot *ja, mutta ja että*).
- Kolligationaalinen valinta: kolligaattien yhtenevä kirjo (kaikilla määritteillä tavataan samat sanaluokka- ja lauseenasemakolligaatit kontekstissa).
- Morfologinen preferenssi: kontekstuaalisen morfologisen preferenssin yhteiset frekventit sananmuodot (<N SG NOM>, <N SG GEN>, <N SG PTV>, <A SG NOM>, <A SG PTV>, <PRON SG NOM>, <V ACT IND PAST SG3>, <V ACT INF1>).
- Semanttinen preferenssi: merkitysluokkien variaatio positiossa O1 suhteellisen samanlainen.

- Semanttinen prosodia: kaikilla astemääritteillä kontekstissaan sekä negatiivista että positiivista prosodiaa osoittavia ominaisuuksia.

Varsin monessa kohtaa analyysin varrella tuli esiin piirteitä, jotka olivat selvästi tyypillisiä jollekin näistä astemääritteistä ja siten osoittivat sen synonyymeistaan poikkeavaksi. Tällaisista piirteistä kokoan seuraavassa kullekin astemääritteelle oman kontekstuaalisen profiilin, johon on kerätty sellaiset ominaisuudet, jotka erottavat määritteen selvästi kahdesta muusta synonyymista. Kontekstuaaliset piirteet on esitetty allekkain astemääritteittäin.

Astemäärite *hyvin*

Kollokationaalinen valinta:

- tilastollisesti merkitsevät kollokaatit *hiljaa, hitaasti, hyvin*, JOKIN, MINÄ, *pian*, PIENI, *varovasti, vähän*, VÄSYNYT
- laaja kollokationaalinen diversiteetti kontekstissa (120 rivin LSS = 60.7)
- omanlaisensa kollokaattien klusteroituminen (tyypillisin <SE OLLA *hyvin*>).

Kolligationaalinen valinta:

- adjektiivien suuri osuus tarkastelualueella (17,2 %)
- omanlaisensa kolligaattien klusteroituminen (tyypillisin 6-jäseninen klusteri <LA PRO VER *hyvin* ADJ LL>).

Morfologinen preferenssi:

- paradigmaattinen preferenssi: *kin*.

Semanttinen preferenssi:

- negatiivinen preferenssi merkitysluokkaan arvo.

Semanttinen prosodia:

- ei selvää prosodiaa, tilannekontekstin merkitys suuri negatiivista prosodiaa tutkittaessa.

Astemäärite *kovin*

Kollokationaalinen valinta:

- frekventit kollokaatit alueella V3–O2 EI, *paljon*
- frekventti kollokaatti EI vasemmanpuoleisessa kontekstissa
- tilastollisesti merkitsevät kollokaatit EI, MONI, *mutta, paljon, TÄRKEÄ, usein*
- alhainen kollokationaalinen diversiteetti vasemmanpuoleisessa kontekstissa, erityisesti positiossa V2
- omanlaisensa kollokaattien klusteroituminen (tyypillisin <EI OLLA *kovin*>)
- kieltöhakuisuus (lauseessa kieltöverbi 67,0 %).

Kolligationaalinen valinta:

- verbikolligaatin suuri osuus (30,0 %) ja konjunktoiden pieni osuus (3,6 %) tarkastelualueella
- omanlaisensa kolligaattien klusteroituminen (tyypillisin 6-jäseninen klusteri <SUB VER VER *kovin* ADJ LL>)

- pitkät kontekstilauseet (sanemäärän ka. 7,28).

Morfologinen preferenssi:

- paradigmaattinen preferenssi: *kaan*
- kontekstuaalinen preferenssi: kieltomuotoiset verbimuodot.

Semanttinen preferenssi:

- positiivinen preferenssi merkitysluokkaan koko/määrä
- negatiivinen preferenssi merkitysluokkaan ulkonäkö.

Semanttinen prosodia:

- selvä negaatioelementin luoma negatiivinen prosodia.

Astemääräte *oikein*

Kollokationaalinen valinta:

- frekventit kollokaatit alueella V3–O2 *hyvin*, HYVÄ
- tilastollisesti merkitsevät kollokaatit *hyvin*, HYVÄ, ISO, *jos*, TUNTUA
- alhainen kollokationaalinen diversiteetti positiossa O1
- omanlaisensa kollokaattien klusteroituminen (tyypillisin <OLLA *oikein* HYVÄ>).

Kolligationaalinen valinta:

- lauseenalkuisuuden suuri osuus tarkastelualueella (11,1 %)
- tyypillisin 6-jäseninen klusteri mahdollisesti <?LA SUB VER *oikein* ADJ SUB/(LL)>
- lyhyet kontekstilauseet (sanemäärän ka. 5,83).

Morfologinen preferenssi:

- negatiivinen preferenssi paradigmaattisessa preferenssissä liitepartikkeleihin.

Semanttinen preferenssi:

- voimakas preferenssi merkitysluokkaan arvo
- negatiivinen preferenssi merkitysluokkiin tuttuus/erilaisuus, aika/ikä.

Semanttinen prosodia:

- selvä positiivinen prosodia.

Astemäärätteiden kontekstuaalisia valintoja esittävistä profiilikuvauksista näkee, että ne ovat varsin erilaisia. Ensinnäkin *hyvin* näyttää olevan eräänlainen yleismääräte, sillä sen valintaa kuvaavat piirteet ovat vähäisemmät kuin muilla. Erityisesti se eroaa kontekstin semanttisilta ominaisuuksiltaan kahdesta synonyymista. Tästä määritteestä poiketen *kovin-* ja *oikein-*määritteillä on vahva omanlaisensa kontekstuaalinen profiili, mikä merkitsee, että niiden käyttötavat ovat *hyvin-*määritettä rajoittuneemmat. Käsillä oleva tutkimus tukee siten vahvasti aiempia havaintoja synonyymisten ilmausten kontekstuaalisesta erilaisuudesta, jota on kuitenkin tähän asti tarkasteltu suurimmaksi osaksi lähinnä kollokationaalisia kytkentöjä analysoimalla (mm. Halliday 1966; Leech 1981; Cruse 1986; Kangasniemi 1997; Jantunen 2001b). Tutkimustulokset vahvistavat myös muut kuin leksikaalisiin myötäesiintymisiin liittyvät kontekstipiirteet selvästi synonyymeja erottaviksi piirteiksi.

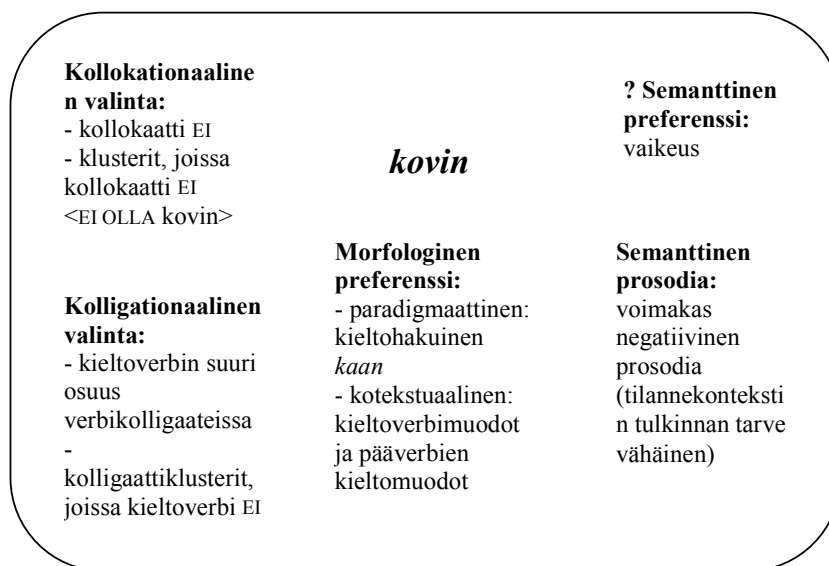
Synonyymit toisistaan erottavat piirteet ovat useassa kohtaa selviä eroja kontekstuaalisten piirteiden osuuksissa. Näin oli etenkin kolligaattianalyysissa, semanttisen preferenssin ja prosodisten piirteiden analyysissa. Sekä koko tarkastelualueetta että pelkästään yhteen positioon keskittyessä löytyi monia selviä tapauksia, joissa jollekin määritteelle oli osoitettavissa muita voimakkaampi preferenssi johonkin kolligaattiin. Erityisesti tämä tuli esiin *kovin*-määritteen preferenssissä verbikolligaattiin sekä koko tarkastelualueetta havainnoitaessa että myös positiokohtaisesti vasemmanpuoleisessa kontekstissa; selitykseksi tälle löytyi *kovin*-määritteen kieltohakuisuus. Myös lauseenasemakolligaattien osuserot osoittautuivat monessa kohdassa synonyymeja erottaviksi muuttujiksi.

Merkityspiirteet määrä, aika/ikä, vaikeus, fyysinen tila/aistimus, tutuus/erilaisuus ja arvo olivat puolestaan sellaisia muuttujia, joiden osuuksien erot tulivat esiin semanttista preferenssiä tarkasteltaessa. Erottavaksi muuttujaksi osoittautui erityisesti *oikein*-määritteen preferenssi arvo-merkityspiirteeseen. Edellisiä ehkä vieläkin selvemmat osuserot oli havaittavissa synonyymien prosodisissa piirteissä, ja synonyymeille oli osoitettavissa omanlaisensa preferenssit negatiivisen ja positiivisen prosodian välillä. Määritteellä *kovin* oli negatiivinen prosodia, *oikein*-määritteellä puolestaan lähes yhtä selvä positiivinen prosodia; *hyvin*-määrite taas oli näiden prosodioiden suhteen lähes neutraali, sillä sen prosodia jakautui negatiivisen ja positiivisen prosodian kesken melkein puoliksi. Erityisesti *hyvin*-määritteen kontekstista ei ollut useinkaan löydettävissä selvää yksittäistä lekseemiä, joka olisi paljastanut prosodian joko negatiiviseksi tai positiiviseksi. Sen sijaan laajemmin tekstiyhteyttä (tilannekontekstia) tarkasteltaessa oli kuitenkin havaittavissa ominaisuuksia, jotka toivat esiin prosodisen ominaisuuden. Toisinaan laajemmankaan kontekstin tarkastelu tai jopa kielenulkoisen kontekstin sisällyttäminen analyysiin ei tuonut selkeyttä *hyvin*-määritteen yksittäisten esiintymien prosodisiin piirteisiin: neutraalien tai epäselvien tapausten määrä olikin selvästi muita suurempi, lähes 8 %.

Vaikka varsin usein synonyymeja erottaviksi piirteiksi osoittautuivat jonkin muuttujan osuserot, tuli tutkimuksessa usein myös vastaan laadullisia ominaisuuksia, jotka olivat synonyymikohtaisia. Erityisen selvästi tällaisiksi osoittautuivat kahden merkitsevyydestin avulla suodatetut tilastollisesti merkitsevät kollokaatit: tutkitut kolme määritettä eivät saaneet yhtään yhteistä tilastollisesti merkitsevää kollokaattia. Tilastollisen testauksen avulla olikin osoitettavissa, että määritteillä on toisistaan poikkeava kollokationaalinen valinta ja että pitkälti yhtenäistävien frekventimpien kollokaattien ohi oli löydettävissä kullekin määritteelle sellaisia kollokaattiosuhteita, joita tavattiin aineistosta useammin kuin olisi voinut olettaa, jos leksikaalinen aineisto jakautuisi tekstissä tasaisesti (mikä on kuitenkin tekstin rakentumisen kannalta sinänsä jo mahdottomuus). Kollokaattien esiintymiselle asetettu vähimmäisvaatimus viidestä esiintymiskerrasta vaikutti jossain määrin analyysiin, sillä tilastollisten ja frekventimpien kollokaattien joukko tuli näin pienemmäksi *kovin*- ja *oikein*-määritteellä kuin niitä paljon taajemmin esiintyvällä *hyvin*-määritteellä. Tässä analyysissa esitettyä tarkemman kuvan voisi saavuttaa suhteuttamalla vähimmäistaajuus ilmauksen esiintymismäärään, jolloin tarkasteluun päätyviä kollokaatteja saataisiin suhteellisesti samankaltainen määrä.

Yksittäisten kollokaattien lisäksi myös klusterirakenteet, sekä leksikaaliset että kolligaattien, paljastuivat erilaisiksi. Määritteet erosivat toisistaan sen mukaan, mitkä olivat tyypillisimpiä leksikaalisia ja kolligaattiklustereita ja mikä oli leksikaalisten klustereiden klusteroitumisaste. Leksikaaliset klusterit erosivat toisistaan myös kokonaisuutena, eivät pelkästään tyypillisimpien klustereiden osalta. Näiden lisäksi myös morfologinen preferenssi erotti synonyymit: kun esimerkiksi liitepartikkeli *kin* oli selvästi *hyvin*-määrittäeseen liittyvä, *kaan* oli puolestaan *kovin*-määrittäeseen, mikä onkin odotettava suffiksoituminen määritteen kieltohakuisuus huomioon ottaen. Synonyymeistaan poiketen *oikein*-määrittäeseen ei suffiksoitunut kumpikaan liitepartikkeleista. Yhdessä aiempien tutkimustulosten valossa näyttää siltä, että morfologinen preferenssi (sekä paradigmaattinen että kontekstuaalinen) on syytä liittää mukaan piirteisiin, joilla leksikaalisten yksiköiden kontekstuaalista profiilia kuvataan.

Vaikka tutkimuksessa analysoitiin kutakin kontekstuaalista muuttujaa omana kokonaisuutenaan, oli monessa yhteydessä havaittavissa, että tietty ominaisuuspiirre kertautui usealla kontekstuaalisella tasolla läpi leksikaalisen yksikön profiilin. Kontekstuaalisia piirteitä ei pidäkään käsitellä toisistaan irrallisina ominaisuuksina, vaan on tarpeen etsiä myös syy-seuraussuhteita eri piirteiden väliltä. Jos prosodiaa pidetään sinclairiläisittäin (Sinclair 1996, 1998) leksikaalisen yksikön uloimpana kerroksena ja kontekstin kielenulkoiseen kontekstiin kytkeväksi piirteeksi, voidaan tässä tapauksessa ajatella, että halutessaan ilmaista jotain negatiivista puhuja valitsee käyttöönsä tähän funktioon sopivat ilmaukset. Kontekstin läpi erottuvasta selvästi kertautuvasta ominaisuudesta käyvät hyviksi esimerkeiksi *kovin*-määritteen kontekstuaalisesta profiilista poimitut kielteisyyteen liittyvät piirteet. Kuten kuviossa 6.8 havaita saattaa, negatio esiintyy määritteen jokaisella tasolla aina konkreettisesta *ei*-kollokaatista abstraktiin negatiiviseen prosodiaan. Semanttinen preferenssi on kuitenkin varustettu kuviossa kysymysmerkillä, koska sitä analysoitiin työssä vain yhdestä positioista. Tässä positiossa ei esiintynyt selvää kieltomerkityspiirrettä (koska se olisi määrittäessään mahdoton), mutta se on toki läsnä määritteen vasemmissa positioissa. Yhdeksi ”negatiiviseksi preferenssiksi” voisi analysista nostaa vaikeus-merkityspiirteen, joka erotti *kovin*-määritteen sen synonyymeista. *kovin* nimittäin määritteli sellaisia vaikeutta merkitseviä leksemejä kuin HANKALA ja VAIKEA, ja lisäksi niissä tapauksissa, missä määrittäksen kohteena oli HELPPO tai YKSINKERTAINEN, helppous ja yksinkertaisuus oli välttävää tai se kiellettiin. Tällaiset havainnot ovat kuitenkin vähäisiä, joskin tarkastelun laajentaminen ohi O1-position nostaisi esiin *kovin*-määritteen semanttisen preferenssin kielteisiä piirteitä ilmaiseviin kielenaineisiin, erityisesti kieltoverbiin.



Kuvio 6.8. *kovin*-määritteen kielteisyyteen liittyvät kontekstuaaliset kytkennät.

7. KONTEKSTUAALISET MERKITYSYKSIKÖT JA SYNONYMIA KÄÄNNÖSSUOMESSA

Luvuissa 3.3.1 ja 4.3. esiteltiin tutkimuksia, joissa on löydetty evidenssiä sekä leksikaalisten yksiköiden epätyypillisistä kombinaatioista että synonyymisten sanojen epätyypillisestä käytöstä käännössuomessa. Seuraavassa analyysissä keskitytään käännössuomessa ilmeneviin kontekstuaalisiin merkitysyksiköihin ja niissä mahdollisesti ilmeneviin epätyypillisiin valintoihin käännössuomessa. Tavoitteena on selvittää edellä tarkasteltuja astemääritteitä ja niiden alkuperäissuomessa tavattavia kontekstuaalisia valintoja hyväksi käyttäen sitä, missä määrin ja millä tavoin käännös- ja alkuperäissuomen kontekstuaaliset merkitysyksiköt poikkeavat toisistaan. Lisäksi käsitellään oletuksia, joita luvussa 5 tehtiin epätyypillisistä frekvensseistä, yksinkertaistumisesta, synonymiasuhteiden muuttumisesta ja lähdekielen vaikutuksesta käännössuomessa. Havaintoyksikköinä käytettäviä astemääritteitä käsitellään yksitellen ja kunkin tapauksen lopussa kuvataan määritteen kontekstuaaliset piirteet ja niiden mahdolliset erot eri aineistoissa. Luvun 6 tapaan analyysi etenee kollokaatioista kolligaatioihin ja morfologiseen preferenssiin ja edelleen semanttiseen preferenssiin ja prosodiaan. Ennen näitä testauksia on kuitenkin ensin syytä tarkastella astemääritteiden esiintymistiheyksiä eri aineistoissa.

7.1. Astemääritteiden esiintymistiheys aineistoissa

Astemääritteiden *hyvin*, *kovin* ja *oikein* esiintymistiheydet on lueteltu taulukkoon 7.1. Testatakseni hypoteesia käännöskielen epätyypillisistä leksikaalisista frekvensseistä olen laskenut astemääritteiden lukumäärän riippuvuussuhteen aineistosta khin neliö -testin (x^2) avulla (testin käytöstä korpustutkimuksessa ks. mm. Oakes 1998).

Taulukko 7.1. Astemääritteiden *hyvin*, *kovin* ja *oikein* esiintymistiheydet aineistoittain.

	AS	KSM	KSE
<i>hyvin</i>	352	709	555
<i>kovin</i>	176	414	306
<i>oikein</i>	121	157	156
Yhteensä	649	1280	1017

Havaittuihin frekvensseihin ja odotusarvoihin perustuva testaus tuottaa x^2 -arvoksi 16,33, mikä ylittää selvästi 99,9 prosentin ($p < 0,001$) luotettavuuden kriittisen arvon, joka on 13,82⁶⁹. Näin ollen astemääritteiden määrä on selvästi riippuvai-

⁶⁹ Raja-arvoista mm. Oakes (1998) sekä Pitkänen ja Kohonen (1984).

nen aineistosta. Jatkotestaus aineistopareittain osoittaa, että määritteiden esiintymistäajuudet ovat aineistoista riippuvaisia vain AS- ja KSM-aineistojen välillä. 95 % luotettavuuden kriittinen arvo on χ^2 -testissä 5,99 (df = 2), joten vain ensimmäinen alla olevista vertailuista ylittää arvon:

AS vs. KSM	$\chi^2 = 16,11$
AS vs. KSE	$\chi^2 = 3,81$
KSM vs. KSE	$\chi^2 = 4,92$

Näin ollen ainoastaan KSM-aineistossa astemääritteiden määrä on tilastollisesti merkittävästi erilainen kuin AS-aineistossa. Käännöskielessä määritteiden määrä näyttää siten olevan suurempi, mutta tämä ei kuitenkaan koske englannista käännettyjä tekstejä, koska kummassakaan vertailuissa, joissa tämä aineisto on mukana, riippuvuutta ei synny. Laskenta on kuitenkin siinä mielessä ongelmallinen, että se ei huomioi osakorpusten suuruuseroa. Vaikka ne ovatkin melko lailla yhdensuuruisia, kannattaa sanemäärien ero ottaa huomioon. Taulukossa 7.2 on astemääritteiden lukumäärät esitetty 10 000 sanetta kohti.

Taulukko 7.2. Astemääritteiden *hyvin*, *kovin* ja *oikein* suhteelliset esiintymistiheydet (f/10 000) aineistoittain.

	AS	KSM	KSE
<i>hyvin</i>	357	660	698
<i>kovin</i>	178	385	385
<i>oikein</i>	123	146	196
Yhteensä	658	1191	1279

Uusi laskenta antaa arvoiksi seuraavat tulokset:

AS vs. KSM	$\chi^2 = 16,04$
AS vs. KSE	$\chi^2 = 4,39$
KSM vs. KSE	$\chi^2 = 5,24$

Khin neliö -testi antaa myös tässä samankaltaisen tuloksen. Tämä tuntuu ristiriitaiselta, koska normitetut frekvenssit ovat molemmissa käännösaineistoissa selvästi suuremmat kuin AS-aineistossa, ja KSE-aineistossa jopa suuremmat kuin KSM-aineistossa. Silti AS- ja KSE-aineistojen ero ei ole testin mukaan merkitsevä. Tämä johtuu siitä, että testi vertaa havaittuja frekvenssejä odotusarvoihin eli arvoihin, jotka olisivat todennäköisiä, jos riippuvuutta ei ole. AS- ja KSM-aineistoja verrattaessa *oikein*-määritteen havaitut ja odotusarvot ovat selvästi erilaisia, mikä tuottaa tilastollisesti merkitsevän tuloksen. AS- ja KSE-aineistojen välillä tämä ero häviää laskukaavan vuoksi, ja astemääritteiden määrä ei ole testin mukaan riippuvainen aineistosta. Näin huolimatta siitä, että normitetut lukumäärät selvästi osoittavat määritteiden olevan yliedustettuja KS-aineistossakin. Vaikka χ^2 -testiä pidetäänkin lingvistin perustyökaluna (Pitkänen ja Kohonen 1984), on siinä myös ongelmansa (ks. mm. Oakes 1998). Tässä tapauksessa ongelmaksi tulee se, että testi ottaa huo-

mioon kaikkien astemääritteiden lukumäärät, jolloin yhden määritteen erot havaituissa ja odotusarvoissa voivat vaikuttaa voimakkaasti tulokseen. Tämän vuoksi jatkan testausta vielä siten, että tarkastelen yksittäisten määritteiden frekvenssejä aineistoissa ja vertailen, miten tulokset suhtautuvat edellä esitettyihin havaintoihin.

Yksittäisten lekseemien yli- ja aliedustumista eri korpuksissa on mahdollista tarkastella Johanssonin (1980) käyttämällä laskentamenetelmällä, jonka avulla hän on vertaillut brittienglannin (Lancaster–Oslo/Bergen Corpus, Johansson ym. 1978) ja amerikanenglannin (Brown Corpus, Kučera ja Francis 1967) sanastoesiintymiä. Kaava on seuraava:

$$\frac{\text{frekvenssi Korpus A} - \text{frekvenssi Korpus B}}{\text{frekvenssi Korpus A} + \text{frekvenssi Korpus B}}$$

Kaavasta saatavan kertoimen arvo vaihtelee $-1:n$ ja $+1:n$ välillä. Positiivinen kerroin ilmoittaa tässä tapauksessa yliedustumista A-korpuksessa (ja aliedustumista B-korpuksessa) ja negatiivinen kerroin vuorostaan yliedustumista B-korpuksessa (ja aliedustumista A:ssa). Kaava edellyttää vertailtavien korpusten samankokoisuutta, joten taulukossa 7.1 esitettyjen, hieman erikokoisiin korpuksiin perustuvien havaittujen arvojen sijaan on alla olevassa vertailussa käytetty korpusten kokoon suhteutettuja arvoja. Näin menetellen kertoimet muodostuvat omassa tutkimusaineistossani korpuspareittain taulukon 7.3 mukaisiksi. Taulukon negatiiviset arvot ilmoittavat yliedustumista vertailuparin jälkimmäiseksi merkityssä aineistossa.

Taulukko 7.3. Astemääritteiden yli- ja aliedustuminen osakorpuksittain.

	AS vs. KSM	AS vs. KSE	KSM vs. KSE
<i>hyvin</i>	-0,30	-0,32	-0,03
<i>kovin</i>	-0,37	-0,37	0,00
<i>oikein</i>	-0,09	-0,23	-0,15

Taulukosta on luettavissa kaksi yleistä linjaa: 1) kaikki astemääritteet yliedustuvat käännessuomessa, ja 2) yliedustuminen ei ole lähdekielestä riippuvaista. Astemääritteiden välillä on kuitenkin jonkinvertaista eroa, sillä *oikein*-määritteen yliedustuminen on hyvin lievää KSM-aineistossa ja KSE-aineistossakin vähäisempää kuin kahdella synonyymilla. Määritteistä selvimmin yliedustuva näyttää olevan *kovin*. Johanssonin (1980: 27–28) tutkimuksessa on suurta yliedustumista osoittaviksi sanoiksi lasketut sellaiset, joiden kerroin asettuu välille $+0,5 - +1,0$ ja $-0,5 - -1,0$. Käsillä olevassa tutkimuksessa yhdessäkään vertailussa kerroin ei ole näiden raja-arvojen välillä, joten määritteiden yliedustumista ei voi pitää kovin huomattavana, mutta kuitenkin selvänä.

Edellä kuvatut testaukset osoittavat ensinnäkin sen, että tarkasteltaessa astemääritteitä ryhmänä niiden esiintymistiheys osoittautuu käännesprosessin tuloksena syntyneessä suomen kielessä suuremmaksi kuin alkuperäissuomessa. Normitettujen frekvenssien ja astemääritteiden keskinäinen vertailu osoittavat, että määritteet yliedustuvat myös englannista käännettyissä teksteissä ja että yliedustuminen voi olla jossain määrin synonyymisidonnaista. Päädynkin nojautumaan jälkimmäisiin ha-

vaintoihin khin neliö -testin ongelmallisuuden vuoksi. Tulos tukee osittain Maura-
sen aiempia (2000b) havaintoja siitä, että synonyymiset sanat ovat eri tavoin prefe-
roituja käänös- ja alkuperäiskielessä. Tässä saatu tulos ei kuitenkaan ole yhtä radi-
kaalisti synonyymeja erotteleva, sillä kaikkien määritteiden frekvenssit osoittivat
ainakin jonkinasteista yliedustumista käänössuomessa, eikä yksikään toisaalta
ollut aliedustunut.

Millainen on sitten kolmen yleisesti esiintyvän astemääritteen suhde toisiinsa?
Onko niistä tyypillisin eli *hyvin* keskinäisessä vertailussa entistäkin tyypillisempi
käänöskielessä eli ”suositaanko” sitä käänöksissä muiden määritteiden kustan-
nuksella? Synonyymisten sanojen frekventeimmän jäsenen suosimistahan on epäil-
lyt muun muassa Kohn (1996). Tätä selvittääkseni olen verrannut määritteiden
osuuksia kaikkien kolmen määritteen yhteisestä esiintymistäajuudesta taulukon 7.2
mukaan. Tulos ei anna viitteitä tyypillisimmän synonyymien suosimisesta. Frekven-
teimmän *hyvin*-määritteen osuus osoittautuu yllättävänkin identtiseksi kaikissa ai-
neistoissa, sillä sen osuus on jokaisessa aineistossa noin 55 % (esim. AS-aineistossa
(357/658)x100). Toiseksi yleisimmän *kovin*-määritteen osuus vaihtelee 27 ja 32
prosentin välillä ja *oikein*-määritteen 12 ja 19 prosentin välillä. Astemääritteiden
keskinäiset osuudet säilyvät vertailun perusteella hyvin samanlaisina aineistoittain,
toisin sanoen käänössuomessa ei yksi synonyymipesyeen jäsenistä osoittaudu
muita korvaavaksi ja suhteellisesti frekventimmin käytetyksi.

7.2. Astemääritteiden kontekstuaalinen valinta käänössuomessa alkuperäis- suomeen verrattuna

7.2.1. *hyvin*

7.2.1.1. Kollokationaalinen valinta

hyvin-määritteen tilastollisesti merkitsevät kollokaatit on lueteltu osakorpuksittain
taulukon 7.4. AS:ssa merkitseviä kollokaatteja on vain 11, mutta KSM:ssa niitä
on 39 ja KSE:ssa 27. Tilastollisesti merkitsevien kollokaattien lukumääräerot eivät
ole kuitenkaan järin erilaisia, kun luvut suhteutetaan aineistoittain vähintään viisi
kertaa esiintyvien kollokaattien määrään, josta tilastollisesti merkitsevät kollokaatit
on laskettu: osuudet ovat aineistoittain 27,5 %, 36,1 % ja 31,8 %. Lievät erot⁷⁰
saattavat johtua astemääritteiden esiintymistäajuuseroista eri korpuksissa.

⁷⁰ Erot eivät ole tilastollisesti merkitseviä: AS vs. KSM $z = 0,9836348$ ($p > 0,05$); AS vs.
KSE $z = 0,48353969$ ($p > 0,05$); KSM vs. KSE $z = 0,631911678$ ($p > 0,05$).

Taulukko 7.4. Tilastollisesti merkitsevät *hyvin*-määritteen kollokaatit osakorpuksittain (vain yhdessä aineistossa merkitsevät kollokaatit kursivoitu; *MI* = *MI*-testi, *t* = *t*-testi).

AS-aineisto			KSM-aineisto			KSE-aineisto		
kollokaatti	<i>MI</i>	<i>t</i>	kollokaatti	<i>MI</i>	<i>t</i>	kollokaatti	<i>MI</i>	<i>t</i>
väsynyt	5,53	2,59	<i>erikoinen</i>	5,07	2,57	yksinkertainen	5,73	3,10
hyvin	4,76	3,97	<i>varhain</i>	5,05	2,38	<i>surullinen</i>	4,94	3,35
hitaasti	4,59	3,32	<i>harvoin</i>	4,96	2,74	väsynyt	4,86	2,55
varovasti	4,44	2,13	yksinkertainen	4,66	2,72	<i>kummallinen</i>	4,86	2,73
<i>hiljaa</i>	3,81	2,46	<i>vaatimaton</i>	4,48	2,14	<i>siisti</i>	4,84	2,16
pian	3,51	2,23	varovasti	4,47	2,34	<i>kalpea</i>	4,81	2,16
vähän	2,79	2,42	<i>vakavasti</i>	4,45	2,34	<i>läheinen</i>	4,78	2,15
pieni	2,00	2,12	<i>selvästi</i>	4,41	2,86	<i>voimakas</i>	4,75	2,36
<i>jokin</i>	1,64	2,45	<i>korkealla</i>	4,34	2,13	ylpeä	4,52	2,34
olla	0,75	5,10	tärkeä	4,34	2,80	<i>onnellinen</i>	4,32	2,85
<i>minä</i>	0,57	2,20	<i>lähellä</i> (adv.)	4,32	2,69	hitaasti	3,95	2,65
			ylpeä	4,32	2,33	hyvin	3,85	4,93
			<i>vaarallinen</i>	4,31	2,69	vaikea	3,68	2,61
			<i>vaalea</i>	4,21	2,50	kaunis	3,67	3,19
			hyvin	4,02	5,78	tärkeä	3,55	2,24
			<i>vakava</i>	3,92	2,29	<i>vihainen</i>	3,50	2,04
			väsynyt	3,88	2,08	nuori	3,16	2,95
			<i>tyytyväinen</i>	3,85	2,08	<i>kylmä</i>	3,15	2,35
			vaikea	3,79	3,21	<i>paha</i>	3,12	2,50
			hitaasti	3,68	2,44	pian	3,02	2,63
			<i>kauas</i>	3,61	2,75	<i>pimeä</i>	2,81	2,10
			<i>sairas</i>	3,60	2,05	<i>silti</i>	2,59	2,21
			<i>vahva</i>	3,52	2,04	<i>pitkä</i>	2,58	3,00
			<i>lyhyt</i>	3,45	2,40	vähän	2,50	2,47
			<i>hieno</i>	3,16	1,99	pieni	2,17	2,80
			<i>lähellä</i> (adp.)	3,13	1,98	<i>näyttää</i>	2,07	3,41
			kaunis	2,92	2,60	olla	0,81	7,42
			<i>nopeasti</i>	2,91	2,29			
			nuori	2,91	3,58			
			pian	2,84	2,72			
			<i>pitkään</i>	2,62	2,05			
			vähän	2,55	3,21			
			<i>vaikuttaa</i>	2,48	2,01			
			<i>ääni</i>	2,28	2,87			
			pieni	1,91	2,65			
			<i>suuri</i>	1,81	2,37			
			<i>paljon</i>	1,46	2,11			
			olla	0,64	6,27			
			<i>ja</i>	0,29	2,32			

Kaikille osakorpuksille yhteisiksi kollokaateiksi erottuvat seuraavat seitsemän, joista OLLA-verbiä lukuun ottamatta kaikki ovat määritteen pääsanoja:

hyvin <*hitaasti, hyvin, OLLA, pian, PIENI, vähän, VÄSYNYT*>.

Tämä seikka korostaa aineistojen keskinäistä samanlaisuutta: nämä kollokaatit esiintyvät määritteen kollokaatteina riippumatta siitä, miten teksti on tuotettu ja mikä lähdekieli on.

Seuraavaksi tarkastelen aineistoja tarkoituksenani löytää sellaiset kollokaatit, jotka ovat tyypillisiä vain käännösuomelle. Sitä ennen kuitenkin lyhyt silmäys alkuperäissuomeen; tästäkin aineistosta löytyy nimittäin kollokaatteja, joita ei tavata muissa korpuksissa. Pelkästään alkuperäissuomelle tyypilliseksi kollokaatioiksi erottuvat *hyvin* <*hiljaa*, JOKIN, MINÄ>. Näistä ovat tilastollisen merkitsevyyden ja myös frekvenssinsä vuoksi mielenkiintoisia kollokaatit JOKIN ja MINÄ. Myöhemmin taulukosta 7.5 (s. 160) on nimittäin havaittavissa, että MINÄ on tavallisin pronominikollokaatti AS-aineistossa, mutta käännöskorpuksissa tyypillisin pronominikollokaatti on HÄN; MINÄ on vasta kolmanneksi yleisin molemmissa käännösaineistoissa. JOKIN on myös frekvenssitilastoissa AS-aineistossa korkeammalla kuin käännösuomen aineistoissa. Frekvenssitilastot ja tilastollinen merkitsevyys tukevat molemmat sitä, että kyseiset kollokaatit ovat tyypillisiä nimenomaan alkuperäissuomelle.

Molemmissa käännösaineistoissa tilastollisesti merkitsevistä kollokaatioista

hyvin <KAUNIS, NUORI, TÄRKEÄ, VAIKEA, YKSINKERTAINEN, YLPEÄ>

yksikään ei ole AS-aineistossa merkitsevien joukossa, joten ne ovat frekvenssiinsä perustuen alkuperäissuomelle epätyypillisiä kollokaatteja. Näistä kuitenkin ainoastaan *hyvin* <YLPEÄ> on sellainen kollokaatio, joka ei esiinny alkuperäissuomen aineistossa ollenkaan, eli se on sekä tilastolliseen testaukseen että havaittuihin esiintymistäajuuksiin perustuen selvästi sellainen kollokaatti, joka on epätyypillinen alkuperäissuomelle (konkordanssi 7.1). Muiden kollokaatioiden esiintymistiheys vaihtelee AS-aineistossa yhdestä esiintymästä viiteen.

Pelkästään KSM-aineistossa esiintyviä tilastollisesti merkitseviä kollokaatteja on runsaasti, yhteensä 25. Näistä suurin osa on O1-positiossa esiintyviä pääsanoja, kuten ERIKOINEN, *nopeasti* ja *paljon* (taulukko 7.4). Mutta kolme muistakin positioista tavattavaa kollokaattia on löydettävissä listalta: *ja*, VAIKUTTAA sekä ÄÄNI. Näistä *ja* on käännösuomen molemmissa aineistoissa selvästi yleisempi kuin alkuperäissuomessa (taulukko 7.5 s. 160), mutta ainoastaan KSM:ssa se on havaintoalueella yleisempi kuin sattuma antaisi olettaa. Erityisen tyypillistä muihin aineistoihin verrattuna on, että tässä aineistossa *ja* esiintyy positiossa V1 ennen määritettä. Tavallisesti *ja* rinnastaa näissä tapauksissa kaksi adjektiivia tai adverbia, joista ainakin jälkimmäinen saa määritteekseen *hyvin*-adverbin (konkordanssin 7.2 rivit 1–8). Muita aineistoja useammin *ja* on KSM:ssa myös positiossa V3, usein aloittamassa lausetta (9–13).

Konkordanssi 7.1. Aineistossa esiintyvät *hyvin* <YLPEÄ> -kollokaatiot.

KSM	inen harvinaisuus, että oli ylpeä siitä ja <i>hyvin</i> tyytyväinen eksoottiseen asemaansa Ei jokainen ole ansioitunut. Minä olen <i>hyvin</i> ylpeä isoisästäni." Fonty yritti hiuk ille viimeinen sakramentti. Isä Mindi oli <i>hyvin</i> ylpeä kutsumuksestaan ja uskoi va lijat. Ja jopa koirankakkakauppias. Olin <i>hyvin</i> ylpeä tiedostani, että sitä omituista vaus. Surustani huolimatta olin isoisästä <i>hyvin</i> ylpeä . Hän oli yksi niitä harvoja i hempieni ajoilta; olin lapsena ollut siitä <i>hyvin</i> ylpeä . Olin pitänyt sitä todisteena r
KSE	kävi kun olin seitsemänvuotias. Olin <i>hyvin</i> ylpeä haavastani, itkin vain aivan temppeleä, ja tšekäläiset ovat siitä <i>hyvin</i> ylpeitä ; mitä pakanallista jumalaa sanoi hän. ”Se oli huikea näytös. Olen <i>hyvin</i> ylpeä siitä, että toin sinut tänne jonka aihe oli raamatullinen. Hän oli <i>hyvin</i> ylpeä tiedoistaan. Minä olin kuitenkin juuri niitä sanoja, muistan sen koska olin <i>hyvin</i> ylpeä . Hän sanoi myös: ”Tästä läh tehnyt hölmön virheen, vaikka oli siitä <i>hyvin</i> ylpeä . Hän vaihtoi talvitakkia erään

Konkordanssi 7.2. Esimerkkejä *ja*-kollokaatista KSM:ssa.

1	asioista ajatteli. Se oli <i>hienopiirteinen ja</i> <i>hyvin</i> <i>herkkä</i> . - Ymmärsin Pietarin puheist
2	nki plus Toompea olisivat sen <i>kininen ja</i> <i>hyvin</i> <i>ilmeikäs</i> runko. Ja Jos saman tien tyy
3	tunut tapaus oli jättänyt <i>salaperäisellä ja</i> <i>hyvin</i> <i>intiimillä</i> tavalla ikuiset jälkensä mei
4	ös pitkä ja hoikka. Hänellä oli <i>kauniit ja</i> <i>hyvin</i> <i>juutalaiset</i> piirteet, hienosti kaareutu
5	otka varmaankin olivat <i>hyvin korkeita ja</i> <i>hyvin</i> <i>vahvoja</i> , mutta jotka siitä katsottuna
6	in kuluneet. Apteekkari puhui <i>tauotta ja</i> <i>hyvin</i> <i>kohteliaasti</i> . ”Olkaa hyvä, tänne päin.
7	ta kai, voihan se olla”, vastasin <i>jurosti ja</i> <i>hyvin</i> <i>naisellisesti</i> . Ja hetken kuluttua: ”Eik
8	Isoäiti poimi lehdet <i>ammattitaitoisesti ja</i> <i>hyvin</i> <i>varovasti</i> , poisti kuihtuneet ja jätti ju
9	Maria istui hetken vaiti LA ja sanoi sitten <i>hyvin</i> varmana: - Hän tiesi sen jo. Leonid
10	lla. Sitä on vaikea ymmärtää, LA ja se on <i>hyvin</i> surullista. Nyt Maria itki. Ilotalon el
11	toimitti sen minulle, LA ja siitä ilmenee <i>hyvin</i> selvästi, millaiseen hirvittävään koe
12	Ja huoneita oli nyt viisi. LA Ja yksi niistä <i>hyvin</i> hämärä, koska ikkuna avautui miltei
13	yn tavaroihin ja rahaan, LA ja tunsit itsesi <i>hyvin</i> mahtavaksi. Näit että muut naiset pe

KSM:ssa jokin myös ”vaikuttaa hyvin joltakin”: VAIKUTTAA esiintyy nimittäin havaintoalueella kuudesti ja on tilastollisesti merkitsevä ainoastaan tässä aineistossa (muissa aineistoissa vain 2 kaksi esiintymää kummassakin). ÄÄNI on edellistä vieläkin tyyppillisempi KSM:n kollokaatti, sillä se esiintyy peräti 13 kertaa ja osoittautuu myös tilastollisesti merkitseväksi (AS:ssa se esiintyy kolmesti ja KSE:ssa vain kahdesti). Kollokaatioanalyysin tässä vaiheessa on syytä muistuttaa kollokaatioanalyysissa käytetystä valintaperusteesta, jonka mukaan kollokaation on esiinnyttävä vähintään kahden eri tekstintuottajan tekstissä (ks. luku 5.4.2). Niinpä sekä VAIKUTTAA- että ÄÄNI-kollokaatit ovat peräisin vähintään kahdesta eri tekstistä, esimerkiksi VAIKUTTAA-tapaukset ovat neljästä eri lähdekieltä edustavasta aineistosta. Myös konkordanssin 7.3 riveillä 13–17 olevat rakenteellisesti samankaltaiset <puhumista ilmaiseva verbi + *hyvin* + adj. + *äänellä*> -tapaukset ovat peräisin eri lähdekielestä käännettyistä teksteistä (lähdekielinä ranska, ruotsi ja viro), eli lähde-tekstejä ja -kieliä on tässäkin tapauksessa useita. KSM-aineistojen kollokaatiot tai niitä laajemmat rakenteet eivät ole siten idiosynkrasian tulosta tai englannista kään-

netyissä teksteissä tavattavia. KSM:n VAIKUTTAA- ja ÄÄNI-tapaukset on lueteltu konkordanssissa 7.3.

Konkordanssi 7.3. VAIKUTTAA- ja ÄÄNI-kollokaatit KSM:ssa.

1	reundlich kumminkin vaikuttaa minusta	<i>hyvinkin</i>	huolestuneelta, ei vain miehensä t
2	tokonekursillaan ja vaikutti muutenkin	<i>hyvin</i>	opinhaluiselta; rullaluistimista tai
3	syksynlehtikasaan. Putoaminen vaikutti	<i>hyvin</i>	hitaalta, minusta tuntui että jouduin
4	mmäistä kertaa kuulutuina ne vaikuttivat	<i>hyvin</i>	karkeilta. Riistäjät. Verenimijät. T
5	oitat, Kostja? - Minusta tämä vaikuttaa	<i>hyvin</i>	oudolta, ystäväiseni! Kukahan tam
6	hdenkymmenen vuoden ikäänsä nähden	<i>hyvin</i>	hienolta vaikuttava musta kävely
7	otti värisevää ääntään esittäen itselleen	<i>hyvin</i>	retorisen kysymyksen: - Sallitaanko
8	lempänä vuorilla. Sekin ääni oli hänelle	<i>hyvin</i>	tuttu. Sitten hän vähitellen tajusi, e
9	frosyne sanoi, mutta hänen äänensä oli	<i>hyvin</i>	hellä. Kun he loikoivat päivällispöy
10	ki hämmästeli hänen äänensä värisevää,	<i>hyvin</i>	kummallista sointia. - Asemalle, to
11	s. Pakkassäällä sellaiset äänet kantavat	<i>hyvinkin</i>	kaus. Kylki kyljessä, seisaallaan ta
12	ta ja vaanimme runoilijan ääntä tuossa	<i>hyvin</i>	vaarallisessa kohdassa. Tsaarin kas
13	iiden kirkkaus sokaisi minut. Tätini sanoi	<i>hyvin</i>	arkisella äänellä kaataessaan höyry
14	aaria - tai vielä suurempi...' Indrek huusi	<i>hyvin</i>	kuuluvalla äänellä : 'Kiitos. Nyt sit
15	vä on sinun luonasi, ura! Avdotja toisti	<i>hyvin</i>	matalalla äänellä kutsuen isoäitiäni
16	eutta pumpuliparran alla. Charlotte sanoi	<i>hyvin</i>	rauhallisella äänellä jänikselle ja o
17	li varomatonta, Maria sanoi, mutta jatkoi	<i>hyvin</i>	päätäväisellä äänellä : - En silti k
18	ko. "Aivan, ymmärrän." Tauko. Ja sitten,	<i>hyvin</i>	samealla äänellä : "Selvä." Hän p
19	arodioi meitä!" Charlotten rauhallisuus ja	<i>hyvin</i>	venäläinen ääni vain pahensivat är

Myös englannista suomennettujen tekstien merkitsevien kollokaattien joukossa on sekä tyypillisesti O1-postiossa olevia kollokaatteja (esim. YKSINKERTAINEN, *pian* ja KYLMÄ) että myös muista asemista tavattavia. Edellä huomattiin KSM:lle tyypilliseksi kollokaatio *hyvin* <VAIKUTTAA>, jota seurasi adjektiivikollokaatti. KSE-aineistosta on löydettävissä samankaltainen myötäesiintymä, jossa kuitenkin VAIKUTTAA-kollokaatin tilalla on NÄYTTÄÄ. Näitä on aineistossa peräti 20, joista osa on esimerkkeinä konkordanssissa 7.4 (rivit 1–5). Toinen vain KSE:sta tavattava merkitsevä kollokaatti, joka ei sijoitu pääsanaposition O1, on *silti* (6–9).

Konkordanssi 7.4. Esimerkkejä NÄYTTÄÄ- ja SILTI-kollokaateista KSE:ssa.

1	ja. Matkan päästä naiset ehkä näyttivät	<i>hyvinkin</i>	eleganteilta sulkineen ja satiineine
2	elehtimällä huoneeseemme. Hän näytti	<i>hyvin</i>	helpottuneelta havaitessaan, että s
3	oikein tajunnut. Kaikki näyttivät olevan	<i>hyvin</i>	kiihtyneessä mielentilassa, mutta en
4	maailmankolkkiin verrattuina. Se näytti	<i>hyvin</i>	kummalliselta, koska aina kun man
5	utamia satasia", sanoi Részsi. Hän näytti	<i>hyvin</i>	surulliselta ja hyvin hauraalta. "Rész
6	olivat yhtä mieltä siitä että ne olivat silti	<i>hyvin</i>	hienoja. Kaikki sanoivat toisilleen:
7	isiä juuri samalla hetkellä, mutta silti oli	<i>hyvin</i>	kaunista. Kaupunki kuulosti vanhu
8	läheiset kuin heillä oli keskenään, mutta	<i>hyvin</i>	läheiset silti . Tässäkään ei ole mit
9	a. Mutta se onnistui häneltä silti : hän joi	<i>hyvin</i>	varovasti puolelta toiselle keinuen.

Taulukossa 7.4 esitettyjä kursivoituja kollokaatteja edelleen analysoitaessa kiinnittyy huomio seikkaan, joka myös erottaa erittäin selvästi kaksi käännösaineistoa toisistaan: kun KSM:ssa O1-kollokaatit ovat lähes yhtä usein adjektiiveja (11 kappaletta) ja adverbeja (9), ja satunnaisesti myös adpositioita (1) ja kvanttoireita (1), KSE:n etupäässä O1-positioissa esiintyvät, vain sille tyypilliset kollokaatit ovat *kaikki* adjektiiveja. Käännösaineistoja eivät siis erota pelkästään yksittäiset tilastollisesti merkitsevät kollokaatit, vaan myös sanaluokat, joita nämä kollokaatit edustavat. *hyvin*-määritteen kontekstissa esiintyviä sanaluokkakollogaatteja käsitellään tarkemmin luvussa 7.2.1.2.

Kollokaattianalyysin lopuksi tarkastelen vielä frekventeimpien kollokaattien järjestystä eri osakorpuksissa. Kaikki astemääritteen *hyvin* saamat vähintään yli viisi kertaa tavatut kollokaatit on lueteltu niiden runsaslukuisuuden vuoksi liitteeseen 3 (taulukko 1) ja tässä yhteydessä on esitetty vain kahdenkymmenen yleisimmän kollokaatin frekvenssit (taulukko 7.5).

Taulukko 7.5. Astemääritteen *hyvin* vähintään 5 kertaa tarkastelualueella V3–O2 esiintyvistä kollokaateista 20 yleisintä. f = frekvenssi, f % = suhteellinen frekvenssi kaikista kollokaateista, F = summafrekvenssi, F % = suhteellinen summafrekvenssi. (Suhteelliset osuudet on laskettu kollokaattien yhteismääristä, jotka ovat AS-aineistossa 1760, KSM-aineistossa 3545 ja KSE-aineistossa 2775.)

	AS-aineisto					KSM-aineisto					KSE-aineisto			
	f	f %	F	F %		f	f %	F	F %		f	f %	F	F %
olla	158	9,0	158	9,0	olla	309	8,7	309	8,7	olla	300	10,8	300	10,8
ja	83	4,7	241	13,7	ja	158	4,5	467	13,2	ja	123	4,4	423	15,2
minä	46	2,6	287	16,3	hän	106	3,0	573	16,2	hän	103	3,7	526	19,0
se	43	2,4	328	18,6	se	87	2,5	660	18,6	se	78	2,8	604	21,8
hän	41	2,3	371	21,1	mutta	44	1,2	704	19,9	minä	48	1,7	652	23,5
hyvin	17	1,0	388	22,0	minä	42	1,2	746	21,0	sanoa	29	1,0	681	24,5
sanoa	17	1,0	405	23,0	hyvin	38	1,1	784	22,1	hyvin	27	1,0	708	25,5
mutta	16	0,9	421	23,9	joka	36	1,0	820	23,1	mutta	25	0,9	733	26,4
että	14	0,8	435	24,7	sanoa	26	0,7	846	23,9	näyttää	20	0,7	753	27,1
jokin	13	0,7	448	25,5	he	18	0,5	864	24,4	me	19	0,7	772	27,8
sinä	13	0,7	461	26,2	ei	17	0,5	881	24,9	tulla	19	0,7	791	28,5
hitaasti	12	0,7	473	26,9	että	17	0,5	898	25,3	että	16	0,6	807	29,1
niin	12	0,7	485	27,6	jokin	17	0,5	915	25,8	joka	16	0,6	823	29,7
tämä	10	0,6	495	28,1	nuori	17	0,5	932	26,3	sinä	15	0,5	838	30,2
ei	9	0,5	504	28,6	vielä	17	0,5	949	26,8	kuin	14	0,5	852	30,7
mies	9	0,5	513	29,1	kaikki	16	0,5	965	27,2	mies	14	0,5	866	31,2
tehdä	9	0,5	522	29,7	tärkeä	16	0,5	981	27,7	pieni	13	0,5	879	31,7
tulla	9	0,5	531	30,2	vähän	15	0,4	996	28,1	pitkä	13	0,5	892	32,1
vain	9	0,5	540	30,7	aika	14	0,4	1010	28,5	kaunis	12	0,4	904	32,6
ihminen	8	0,5	548	31,1	me	14	0,4	1024	28,9	surullinen	12	0,4	916	33,0
joka	8	0,5	556	31,6										
pieni	8	0,5	564	32,0										
vähän	8	0,5	572	32,5										

Yleisimpiä kollokaatteja ovat odotetusti frekventit sanat OLLA sekä *ja* jokaisessa aineistossa. Niiden jälkeen seuraa pronominikollokaatteja (MINÄ, HÄN, SE), joiden järjestys kuitenkin on erilainen alkuperäissuomen aineistossa käännöskorpuksiin

verrattuna. Yleisimpien kollokaattien joukkoon mahtuvat jokaisessa aineistossa myös *hyvin, mutta* ja SANOA.

Myös kollokaattien klusteroitumista tarkastelemalla voidaan tutkia edelleen mahdollisia käänköskielelle tyypillisiä ja siten alkuperäissuomelle epätyypillisiä leksikaalisia kombinaatioita. Keskityn analyysissä vähintään neljä kollokaattia sisältäviin lemmaklustereihin, ja kolmisanaisista klustereista vain kolmeen yleisimpään, koska kaikkien kolmisanaisien klustereiden joukko on hyvin runsas analysoitavaksi. Kuten luvussa 6, myös tässä käänkössuomen analyysiin mukaan otetut klusterit ovat vähintään kahden eri kielenkäyttäjän tuottamia. Klusterit on lueteltu aineistoittain taulukossa 7.6.

Taulukko 7.6. *hyvin*-määritteen klusteroituminen aineistoittain.

	sanoja	V3	V2	V1	noodi	O1	O2	f
AS	4	että	minä	olla	<i>hyvin</i>			2
	3		se	olla	<i>hyvin</i>			13
			hän	olla	<i>hyvin</i>			11
			minä	olla	<i>hyvin</i>			11
KSM	4	minä	olla	olla	<i>hyvin</i>			2
		ja	hän	olla	<i>hyvin</i>			2
		hän	olla	vielä	<i>hyvin</i>			2
		hän	olla	olla	<i>hyvin</i>			2
			minä	olla	<i>hyvin</i>	vaikea		3
			se	olla	<i>hyvin</i>	vaikea		2
			olla	vielä	<i>hyvin</i>	nuori		2
	3		hän	olla	<i>hyvin</i>			19
			olla	olla	<i>hyvin</i>			13
			se	olla	<i>hyvin</i>			12
KSE	5		hän	olla	<i>hyvin</i>	pitkä	hän	2
	4	minä	se	olla	<i>hyvin</i>			3
		se	oli	jokin	<i>hyvin</i>			2
			hän	olla	<i>hyvin</i>	pitkä		4
			ja	olla	<i>hyvin</i>	vihainen		2
			päättää	nousta	<i>hyvin</i>	varhain		2
			olla	se	<i>hyvin</i>	ylpeä		2
				olla	<i>hyvin</i>	erilainen	kuin	2
				olla	<i>hyvin</i>	kaunis	ja	2
	3		se	olla	<i>hyvin</i>			21
			hän	olla	<i>hyvin</i>			12
			minä	olla	<i>hyvin</i>			9

Analyyisin perusteella *hyvin*-määritteen klusteroituminen on varsin erilaista eri aineistoissa. Ensinnäkin alkuperäissuomen aineistossa nelisanaisia klustereita on vain yksi, mutta käänkösaaineistoissa näitä on useita. Määritteen klusteroituminen nelisanaisiin klustereihin osoittautuu myös varsin eriaistiseksi: AS:ssa klusteroitumisaste on 0,6, KSM:ssa 2,1 ja KSE:ssa vieläkin suurempi, 3,4. Ero KSE- ja AS-aineiston välillä on merkitsevä ja KSM- ja AS-aineiston välillä suuntaa antava⁷¹, eli

⁷¹ KSE vs. AS $z = 2,786364$ ($p < 0,01$); KSM vs. AS $z = 1,890170$ ($p < 0,10$).

käännössuomessa – erityisesti englannista käännettyissä teksteissä – *hyvin*-määrittelyllä on taipumus klusteroitua alkuperäisruomea useammin nelisanaisiin klustereihin. Tätä tukee myös se, että ainoa viisisanainen klusteri löytyy englannista käännettyistä teksteistä. Tulosta voidaan pitää myös jonkinlaisena tukena oletukselle leksikaalisten kombinaatioiden yksinkertaistumisesta, sillä mitä vakiintuneemmissa klustereissa lekseemi esiintyy, sitä vähemmän lekseimin kontekstissa on tilaa leksikaaliselle variaatiolle.

Nelisanaiset klusterit osoittautuvat myös rakenteeltaan melko erilaisiksi: AS-aineiston ainoa tällainen klusteri on <että MINÄ OLLA *hyvin*>, KSM-aineistossa tavallisin on <MINÄ OLLA *hyvin* VAIKEA> ja KSE:ssä vuorostaan <HÄN OLLA *hyvin* PITKÄ>. Nelisanaisien klustereiden perusteella voi tehdä päätelmiä erilaisesta klusteroitumisen suunnasta käännösaaineistojen kesken (koska AS-aineistossa klustereita on vain yksi, on tämän klusterin osoittamaa klusteroitumissuuntaa pidettävä vain viitteellisenä). KSM-aineistossa klusteroituminen näyttää olevan selvästi vasemmalle painottuvaa, sillä klusterit yltyvät useasti V3-positioon asti. KSE-aineistossa klusteroituminen näyttää painottuvan enemmänkin oikealle: O1-positiossa on runsaasti klustereihin kuuluvia lekseemejä, ja 2 klusteria yltyä jopa positioon O2 asti.

Kolmiosaiset taajimmin esiintyvät klusterit ovat rakenteeltaan ja leksikoltaan jokseenkin samanlaisia kaikissa aineistoissa. Merkille pantavaa on kuitenkin, että alkuperäisruomessa kaikki kolme frekventeintä klusteria ovat lähes yhtä yleisiä, mutta molemmista käännösaaineistoista erottuu yksi kolmisanainen klusteri muita selvästi tavallisemmaksi, mikä saattaa olla osoitus leksikaalisten kombinaatioiden yksinkertaistumisesta käänössuomessa. Klusteroitumisaste kolmisanaisiin klustereihin on seuraavanlainen: AS 23,0, KSM 22,7 ja KSE 30,3. Tulos tukee edellä esiin tullutta englannista käännettyjen tekstien klusteroitumisen voimakkuutta (erot KSE-aineiston ja molempien muiden aineistojen välillä ovat merkitseviä⁷²), mutta ei kuitenkaan sitä, että käänöksissä olisi yleensä klusteroituminen voimakkaampaa, koska AS- ja KSM-aineistojen arvot ovat liki samat.

Klusterianalyysi paljasti käänössuomesta ainakin jossain määrin alkuperäisruomelle epätyypillisiä kombinaatioita ja erityisesti osoittaa käännösaaineistojen pitkät klusterit erilaisiksi, mikä tukee kollokaatioanalyysistä saatuja tuloksia. Lisäksi klusterianalyysi näyttää tuottavan myös tuloksia, jotka viittaavat käännöskielen leksikaalisten kombinaatioiden yksinkertaistumiseen. Tätä osoittivat erilainen taipumus pitkiin klustereihin ja yleisimmän kolmisanaisen klusterin suuret määrät käännösaaineistoissa. Suomennosten lähdekielellä näyttää kuitenkin olevan vaikutusta, sillä englannista suomennetuissa teksteissä klusteroituminen on voimakkainta. Yksinkertaistumishypoteesia on kuitenkin syytä tarkastella vielä yksityiskohtaisemminkin. Tässä käytän mittareina frekventeimpien kollokaatioiden osuutta aineistoissa sekä kollokationaalista diversiteettiä mittaavaa lemma–sane-suhdetta (LSS).

Laviosan tulokset osoittivat, että englannin 200 tavallisinta sanaa peittävät koko aineiston sanastosta suuremman osan käänöksissä kuin alkuperäiskielisissä teksteissä (Laviosa-Braithwaite 1996: 128, 132). Samansuuntaisiin tuloksiin on hiljattain päätyntä myös Nevalainen (tulossa), joka on tarkastellut sadan ja tuhannen yleisimmän sanamuodon ja lemmän osuuksia suomessa: käänössuomessa taa-

⁷² KSE vs. AS $z = 2,387152$ ($p < 0,05$); KSM vs. AS $z = 3,040810$ ($p < 0,01$).

jimmin esiintyvät lekseemit kattavat suuremman osan kuin alkuperäissuomessa. Jos koko aineiston leksikko on osoitettavissa yksinkertaisemmaksi käännskielessä, voitaneen olettaa, että sanastollista yksinkertaistumista voisi olla havaittavissa myös lekseemien koteksteissa. Oletuksen voinee tehdä etenkin silloin, kun tarkastelualue on niin laaja, että siihen mahtuu sekä kielen tavallisinta sanastoa että myös harvakseltaan esiintyviä lekseemejä. Näin on tässä tutkimuksessa, kuten aiemmin esitetyistä monista taulukoista on havaittavissa: tarkastelualueella esiintyy kielen frekventeintä ydinsanastoa mutta myös määritteiden koteksteille tyypillistä sanastoa. Määritteiden kotekstien kollokaattien frekvenssilista antaa siten mahdollisuuden tarkastella, ovatko niiden kotekstit sanastollisesti yksinkertaistuneita käännskiosuomessa – tuloksia ei voi luonnollisestikaan Laviosan ja Nevalaisen tavoin yleistää koko aineiston sanastoa koskevaksi.

hyvin-määritteen kotekstin kollokaattien osuudet on merkitty taulukkoon 7.7. Sekä viiden että kymmenen yleisimmän kollokaatin osuudet ovat KSE-aineistossa suuremmat kuin yleensä käännskiosuomessa, mutta eivät tilastollisesti merkitsevästi suuremmat kuin alkuperäissuomessa. Kuitenkin vertailussa AS-aineistoon ylittyä 90 prosentin merkitsevyystaso molemmissa tapauksissa ($z = 1,90$; $z = 1,70 > z_{.10} = 1,64$), joten KSE-aineistojen suurempaa osuutta suhteessa AS-aineistoon voidaan pitää suuntaa antavana. Yleisimpien kollokaattien suurempi osuus englannista käännettyissä teksteissä osoittaa siis jonkinlaista leksikaalista yksinkertaistumista, sillä niissä kotekstin aivan yleisimmät lekseemit ovat taajemmin käytettyjä kuin yleensä käännskiosuomessa. KSE-aineiston leksikaalista yksinkertaistumista tukevat myös yhteneväistä suuntaa osoittavat arvot suhteessa alkuperäissuomeen. Monilähdekielisessä aineistossa vastaavaa frekventeimpien kollokaattien yliedustumista ei kuitenkaan ole havaittavissa. Näin ollen Laviosan ja Nevalaisen havainnot saavat tukea vain englannista käännetystä aineistosta, mikä tarkoittaa, että lähdekieli voi tässä yhteydessä vaikuttaa tuloksiin.

Taulukko 7.7. *hyvin*: yleisimpien kollokaattien summafrekvenssien suhteellisten osuuksien vertailu.

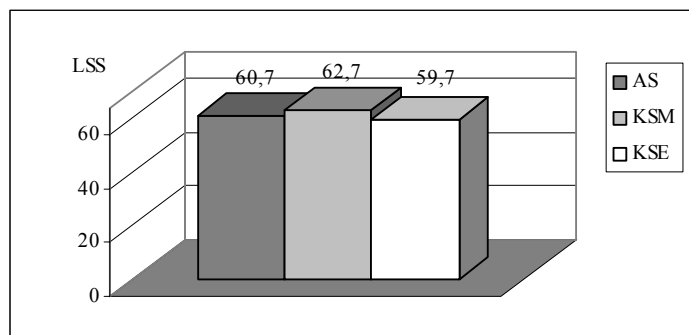
	AS vs. KSM	AS vs. KSE	KSM vs. KSE
5 yleisimmän osuus	21,1 % vs. 19,9 % $z = 1,04, p > 0,05$	21,1 % vs. 23,5 % $z = 1,90, p > 0,05$	19,9 % vs. 23,5 % $z = 3,49, p < 0,001$
10 yleisimmän osuus ^o	26,2 % vs. 24,4 % $z = 1,44, p > 0,05$	26,2 % vs. 28,5 % $z = 1,70, p > 0,05$	24,4 % vs. 28,5 % $z = 3,71, p < 0,001$

^oKymmenen yleisimmän kollokaatin joukkoon on laskettu AS:ssa ja KSE:ssa 11 kollokaattia, koska rajakohdan ylittävillä kollokaateilla on sama frekvenssi.

Edellä luvussa 6 huomattiin sekä koko tarkastelualueen että positiokohtaisen kollokationaalisen diversiteetin erottelevan synonyymisia määritteitä toisistaan. Diversiteetin mittauksessa käytettyä lemma–sane-suhdetta (LLS) voidaan soveltaa hyvin myös eri aineistoissa esiintyvien määritteiden kotekstin leksikaalisen rikkauksen mittaamiseen. Kuva kotekstien leksikon yksinkertaistumisesta tulee myös edellistä analyysia tarkemmaksi, koska nyt mitataan kotekstin koko sanaston rikkautta. Tarkastelualueen leksikaalista diversiteettiä kuvaavat tulokset on esitetty kuviossa

7.1 ja positiokohtaista diversiteettiä kuviossa 7.2. Molemmat suhdeluvut on laskettu noin 120 konkordanssiriviesiintymän avulla kaikissa aineistoissa (vrt. kuvio 6.1).

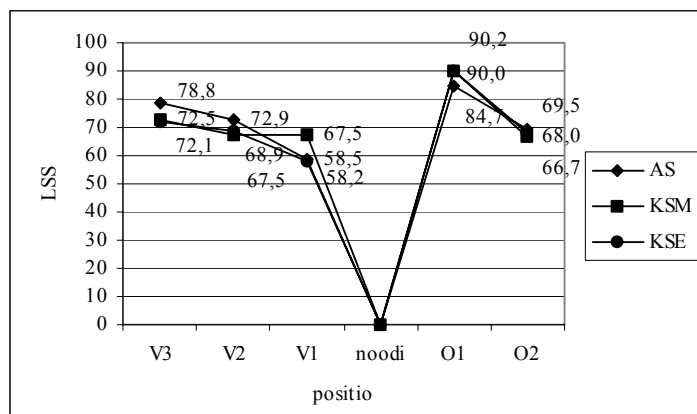
Koko tarkastelualueelta V3–O2 mitattu kollokationaalinen diversiteetti ei anna todisteita kotekstin sanaston yksinkertaistumisesta, sillä KSM-aineistossa LSS on jopa hiiven suurempi kuin alkuperäissuomessa ja englannista suomennetuissa teksteissä likipitään yhtä suuri kuin alkuperäissuomessa. Erot eivät ole tilastollisesti merkitseviä⁷³. Positiokohtaisia LSS:ita kuvaavat käyrät osoittautuvat kaikissa aineistoissa hyvin samanlaisiksi. Suurimmillaan ero on positiossa V1 (9,3 prosenttiyksikköä), mutta siinäkin asemassa aineistojen väliset erot eivät ole tilastollisesti merkitseviä⁷⁴. *hyvin*-määritteen kollokationaalinen diversiteetti ei siten ole erilainen aineistojen välillä. Näin ollen koko tarkastelualueelta laskettu diversiteetti ei tue leksikaalista yksinkertaistumista, vaikka tätä tukevia havaintoja löytyi taajimmin esiintyvien sanojen ja klustereiden analysista.



Kuvio 7.1. *hyvin*: lemma–sane-suhde tarkastelualueella V3–O2.

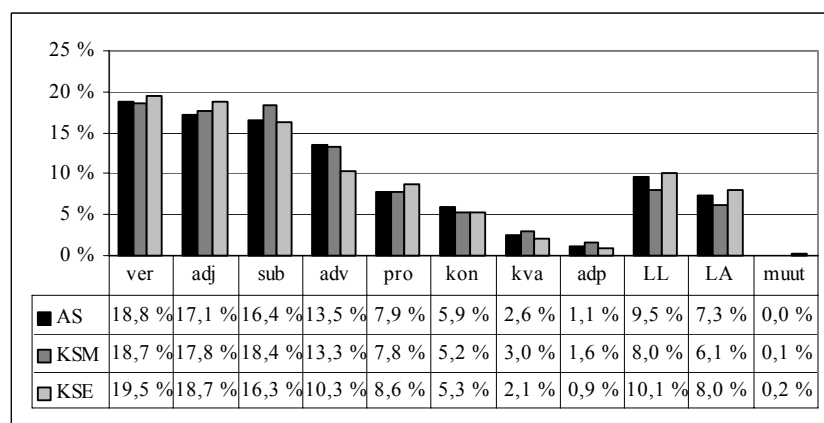
⁷³ AS vs. KSM $z = 0,731914$ ($p > 0,05$); AS vs. KSE $z = 0,355131$ ($p > 0,05$); KSM vs. KSE $z = 1,102826$ ($p > 0,05$).

⁷⁴ AS vs. KSM $z = 1,442118$ ($p > 0,05$); AS vs. KSE $z = 0,022104$ ($p > 0,05$); KSM vs. KSE $z = 1,469992$ ($p > 0,05$).

Kuvio 7.2. *hyvin*: lemma-sane-suhde positioittain.

7.2.1.2. Kolligatiivinen valinta

Seuraavaksi siirrän tarkastelun *hyvin*-määritteen kolligatiiviseen valintaan, jonka avulla pyrin tutkimaan edelleen käännöskielen epätyypillisiä kombinaatioita ja yksinkertaistumista. Tarkastelu koostuu kolligatiivien osuuksien, kolligatiivien klusteroitumisen ja lausepituuden analysoinnista. *hyvin*-määritteen tarkastelualueella V3–O2 olevien sanaluokkakolligatiivien osuudet on esitetty kuviossa 7.3, ja ne kolligatiivit, joiden osuuksien erot ovat aineistojen välillä tilastollisesti merkitseviä, on lueteltu taulukossa 7.8.

Kuvio 7.3. *hyvin*: kolligatiivit koko tarkastelualueella V3–O2.

Taulukko 7.8. Kolligaatit, joiden osuudet eroavat toisistaan tilastollisin perustein aineistojen välillä.

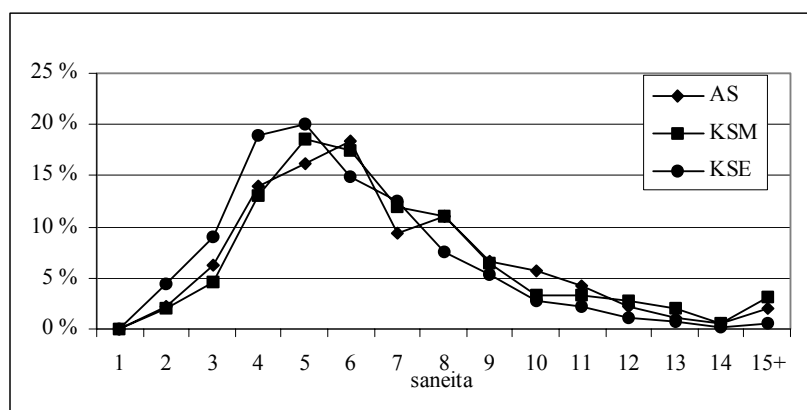
AS - KSM	AS - KSE	KSM - KSE
		adpositiot z = 2,30, p < 0,05
		adverbit z = 3,58, p < 0,001
		kvanttorit z = 2,17, p < 0,05
–	adverbit z = 3,21, p < 0,01	substantiivit z = 2,22, p < 0,05
		lauseenalkuisuus z = 2,87, p < 0,01
		lauseenloppuisuus z = 2,88, p < 0,01

Kuviosta 7.3 on luettavissa, että määritteen sanaluokkakolligaatit ovat samoja kaikissa aineistoissa: yhdessäkään aineistossa ei ole sellaista sanaluokkakolligaattia, jota ei olisi jossakin toisessa aineistossa, toisaalta myöskään yksikään sanaluokkakolligaatti ei puutu aineistoista. Näin ollen sanaluokkakolligaattianalyysi ei tuo esille epätyypillisiä kombinaatioita. Sen sijaan kolligaattien osuuksissa on jonkin verran eroja aineistoittain, joten tästä saadaan tukea oletukselle epätyypillisistä frekvensseistä. Alkuperäissuomen ja käännössuomen aineistojen kolligaattien osuuksien erot ovat kuitenkin olemattomat: AS- ja KSM-aineistoja ei erota yhdenkään kolligaatin osuus, ja ainoastaan adverbikolligaatin osuus englannista käännetyissä teksteissä on tilastollisesti merkitsevästi erilainen kuin alkuperäissuomessa (taulukko 7.8).

Aineiston lähempi tarkastelu paljastaa, että adverbikolligaatin vähäinen osuus KSE-aineistossa johtuu adverbikolligaattien pienestä määrästä *hyvin*-määritteen vieruspositioissa. Kolligaatin osuus on AS:ssa 27,4 % kaikista O1-kolligaateista, KSE:ssa osuus on 20,7 %, ja ero on myös tilastollisesti merkitsevä; määritettä edeltävässä V1-positiossa osuus on AS:ssa 17,9 % ja KSE:ssa 13,3, ja ero on suuntaa antava.⁷⁵ Tilanne on samanlainen käännösaineistojen kesken, sillä niidenkin välillä adverbikolligaattien osuus on yksi erottava tekijä (nimenomaan pääsanapositiiossa), ja ero on erittäin merkitsevä. Adverbien vähäinen määrä tuli jo osittain esille kollokaatioanalyysissä, missä huomattiin, että englannista käännetyissä teksteissä *hyvin*-määritteen tilastollisesti merkitsevä kollokaattina oli harvoin adverbi (taulukko 7.4). Käännösaineistot eroavat toisistaan kolligaattien osuuksien perusteella muutenkin erittäin selvästi, sillä peräti kuuden kolligaatin osuudet ovat merkitsevästi erilaisia aineistojen välillä. KSE-aineistossa niin substantiivi-, kvanttori-, adpositio- kuin jo edellä mainitun adverbikolligaatinkin osuus on merkitsevästi pienempi kuin käännöksissä yleensä, mutta lauseenalkuisuuden ja -loppuisuuden osuudet taas merkitsevästi suurempia (edellä kuvio 7.3 ja taulukko 7.8).

⁷⁵ O1-positiossa AS vs. KSE z = 2,465150 (p < 0,05); V1-positiossa AS vs. KSE z = 1,870678 (p < 0,10).

Lauseenalku- ja lauseenloppukolligaattien erilainen osuus käänösaineistossa antaa olettaa, että myös lauseiden pituus olisi aineistossa erilainen, tässä tapauksessa lausepituuden voisi olettaa olevan KSE-aineistossa pienempi kuin muissa aineistoissa. Tätä tarkastellaan kuvion 7.4 ja taulukon 7.9 avulla.



Kuvio 7.4. *hyvin*: lauseiden sanemäärät.

Taulukko 7.9. *hyvin*: lauseiden sanemäärien keskiarvot ja hajonta sekä keskiarvojen erojen tilastollinen merkitsevyys aineistoittain.

	AS	KSM	KSE
keskiarvo (saneita)	6,68	6,90	5,83
vaihteluväli (saneita)	14 [2–16]	21 [2–23]	20 [2–22]
	AS–KSM	AS–KSE	KSM–KSE
merkitsevyys	$z = 1,10$ ($p > 0,05$)	$z = 4,70$ ($p < 0,001$)	$z = 6,72$ ($p < 0,001$)

Tyypillisin lausepituus vaihtelee alkuperäisruomen ja käänössuomen aineistojen välillä: AS-aineistossa lause on tyypillisimmin kuusisanainen ($n = 65$), käänösaineistoissa viisisanainen (KSM $n = 131$, KSE $n = 111$). Nelisanaisten lauseiden osuus on KSE-aineistossa selvästi suurempi kuin muissa aineistoissa ja kuusisanaisten vuorostaan pienempi (kuvio 7.4). Tämä viittaa englannista suomennettujen tekstien lyhyempään kotekstuaaliseen lausepituuteen, mikä näkyikin selvästi pienempänä lausepituuden keskiarvona (taulukko 7.9). Vaikka yleensä käänöskielen lauseet näyttävät olevan hivenen pitempiä kuin alkuperäisruomen, ero ei ole merkitsevää. KSE-aineiston lausepituuden ero on sen sijaan erittäin merkitsevä suhteessa molempiin muihin aineistoihin. Englannista käännettyissä teksteissä *hyvin*-määräte esiintyy siten lyhyemmissä lauseissa kuin mitä käänöskielellä yleensä ja AS-aineistossa. Laviosa (Laviosa-Braithwaite 1996: 141) puolestaan havaitsi, että käännettyissä kaunokirjallisissa teksteissä lauseet olivat merkitsevästi pitempiä kuin alkuperäisaineistossa. Laviosan tulos ei siten tukenut hänen oletustaan käänöskielen tyyllillisestä yksinkertaistumisesta, kuten ei tässäkään työssä AS- ja KSM-aineisto-

jen vertailu. Kaunokirjallisuuden osalta Laviosa ei pystynyt tutkimaan lähdekielen vaikutusta aineiston vähyiden vuoksi (mts. 146). Omassa aineistossani nimenomaan englannista käännettyt tekstit paljastuvat erilaisiksi, sillä KSE-aineiston kontekstilauseet ovat keskimäärin huomattavasti lyhyempiä kuin yleensä käännskielellä.⁷⁶

Viimeinen kolligaattianalyysi koskee kolligaattien klusteroitumista. AS-aineistossa tyypillisin *hyvin*-määritteen kolligaattiklusteri on

<LA PRO VER *hyvin* ADJ LL>,

joka kattaa 4,3 prosenttia kaikista esiintymistä (luku 6.2.3). Tyypillisin klusteri on sama sekä KSM- että KSE-aineistossa, eli kolligaattien klusteroituminen ei tue oletusta epätyypillistä kombinaatioista käännskielellä. Klusterin osuus on KSM:ssä 1,8 % (n = 13) ja KSE:ssä 5,0 % (n = 28). KSM:n osuuden vähäisyys on muihin aineistoihin verrattuna tilastollisesti merkitsevä⁷⁷, mikä tarkoittaa sitä, että yleisin klusteri kattaa yleensä käännskielellä pienemmän osuuden kaikista klustereista kuin alkuperäiskielisissä ja englannista suomennetuissa teksteissä, mikä viittaa suurempaan variaatioon aineistossa, jossa lähdekieliä on useita. Kolligaattiklustereiden analyysi ei siten tue oletusta käännskielellä yksinkertaistumisesta, mutta tuo jälleen esiin sen, että englannista käännettyjen tekstien kieli voi poiketa ominaisuuksiltaan käännskielestä, jossa on monien lähdekielten vaikutusta.

7.2.1.3. Morfologinen preferenssi

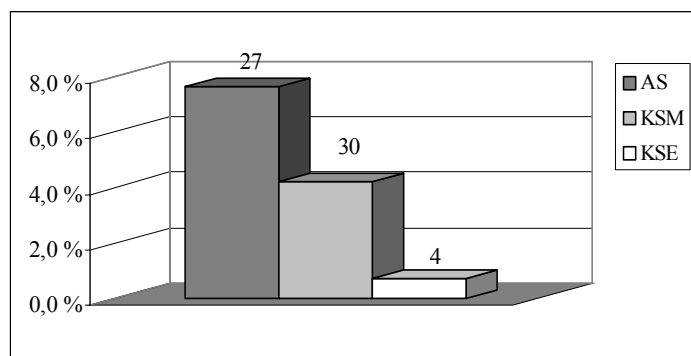
Morfologisesta preferenssistä on käännskielellä kannalta erityisen mielenkiintoinen paradigmaattinen preferenssi, joka tarkasteltavien lekseemien kohdalla tulee ilmi liitepartikkelien suffiksoitumisena määritteisiin. Käännskielellä osalta keskitynkin vain paradigmaattiseen morfologiseen preferenssiin, koska se tarjoaa mielenkiintoisia yhtymäkohtia aikaisempaan käännskielellä tutkimukseen. Viimeaikaisimpia käännskielellä universaalihypoteeseja on nimittäin Sonja Tirkkonen-Conditiin (2002, 2004) postulaatti niin sanotusta kohdekielispesifisten uniikkiaineiden ali-edustumisesta käännskielellä. Tirkkonen-Conditiin hypoteesi perustuu Reissin (1971) esittämään oletukseen, jonka mukaan käännskielellä olisi vähemmän sellaisia kielenaineita, joille ei ole löydettävissä yksiselitteistä käännskielellä vastinetta muista kielistä. Käännskielellä prosessissa tämä tarkoittaisi sitä, että kääntäjä ei tulisi käyttäneeksi kohdekielispesifisiä vastineettomia ilmauksia, koska niille ei olisi selkeää stimulusta lähdetekstissä ja -kielessä. Tätä tukevia havaintoja onkin saatu jo muutamassa aiemmassa tutkimuksessa. Ensinnäkin Tirkkonen-Conditiin (2000b, 2001) on huomannut, että modaalistyypilliset verbit kuten *hennoa*, *iljetä*, *jaksaa*, *joutua*, *maltaa*,

⁷⁶ Laviosa (Laviosa-Braithwaite 1996: 129) havaitsi kuitenkin, että italiasta ja espanjasta englanninnettujen tekstien lausepituus oli pitempi kuin saksasta ja ranskasta käännettyjen, mikä tukee lähdekielen vaikutusta.

⁷⁷ KSM vs. AS $z = 2,322924$ ($p < 0,05$); KSM vs. KSE $z = 3,198473$ ($p < 0,01$).

uskaltaa, *viitsiä* aliedustuvat suomennoksissa, koska niille ei ole selvää leksikaalisen vastinetta lähdekielessä (Tirkkonen-Conditiin tutkimuksessa englannissa). Vastaavankaltaisia havaintoja on tehnyt myös Kujamäki (2004) tarkastellessaan reaaliota *keli*, *kinos* ja *hanki*. Eskola (2002) teki puolestaan samansuuntaisia havaintoja syntaktisista rakenteista: hänenkin syntetisoivia rakenteita koskevassa tutkimuksessaan ilmeni, että lähdekielen stimulus vaikuttaa selvästi kohdekielispesifisten aineiden esiintymiseen.

Oman tutkimukseni kannalta ovat merkittäviä liitepartikkelit, jotka ovat suomen kielelle tyypillisiä kielenaineksia, mutta joita ei esiinny korpuksen indoeurooppalaisissa lähdekielissä, joskin kylläkin virossa ja unkarissa.⁷⁸ Liitepartikkelit on myös Tirkkonen-Conditi (2000a: 147) maininnut vastineettomiksi kielenaineksiksi, jotka voivat aliedustua käännösuomessa, ja myöhemmässä tutkimuksessaan (2004) hän on saanut selvää evidenssiä liitepartikkelien *kin* ja *han* aliedustumisesta käännösuomessa. Luvussa 6.3 ilmeni, että määritteillä on nimenomaan liitepartikkeleihin liittyvää morfologista preferenssiä. Onkin hedelmällistä tarkastella, miten liitepartikkelit edustuvat käännösuomessa.



Kuvio 7.5. *hyvin*: liitepartikkelien (*kin*) osuus aineistoittain.

hyvin-määritteen liitepartikkelillisten muotojen osuudet on kuvattu aineistoittain kuviossa 7.5. Muotojen osuus on selvästi erilainen aineistoittain: alkuperäis-suomessa liitepartikkelillisiä muotoja on kaikista määritteen esiintymistä 7,7 %, KSM-aineistossa 4,2 % ja englannista käännettyissä teksteissä vaivaiset 0,7 %. Erot ovat kaikkien korpusten välillä tilastollisesti merkitseviä⁷⁹. Tämä tarkoittaa, että käännösuomessa liitepartikkelimuodot aliedustuvat lähdekielestä riippumatta, mikä tukee Tirkkonen-Conditiin hypoteesia. Lähdekieli näyttää kuitenkin vaikuttavan esiintymisiin, sillä englannista suomennetuissa teksteissä osuus on pienempi kuin käännöksissä yleensä. Syytä pieneen frekvenssiin voisi etsiä englannin systeemistä,

⁷⁸ Kiitän professori Magdolna Kovácsia (HY) ja professori Karl Pajusalua (HY) unkarin ja viron kielen liitepartikkelijärjestelmää koskevista tiedoista.

⁷⁹ AS vs. KSM $z = 2,339365$ ($p < 0,05$); AS vs. KSE $z = 5,613797$ ($p < 0,001$); KSM vs. KSE $z = 3,828524$ ($p < 0,01$).

josta ei löydy vastaavaa liitepartikkelirakennetta. Jos lähdekielen stimulus vaikuttaisi näin selvästi, olisi luultavaa, että KSM-korpuksen esiintymät olisivat peräisin lähinnä virosta, jossa on suomen *kin*-liitettä vastaava liitepartikkeli (*gi*), joka voi yhdistyä astemääritteisiin. KSM-aineiston *kin*-liitteiset tapaukset eivät ole kuitenkaan peräisin pelkästään virosta suomennetuista teksteistä, vaan niitä on yhtä hyvin indoeurooppalaisista kielistä tehdyissä suomennoksissa. Lähdekielen stimulus ei siis ole *kin*-liitepartikkelien esiintymisen selitys.

7.2.1.4. Semanttinen preferenssi ja prosodia

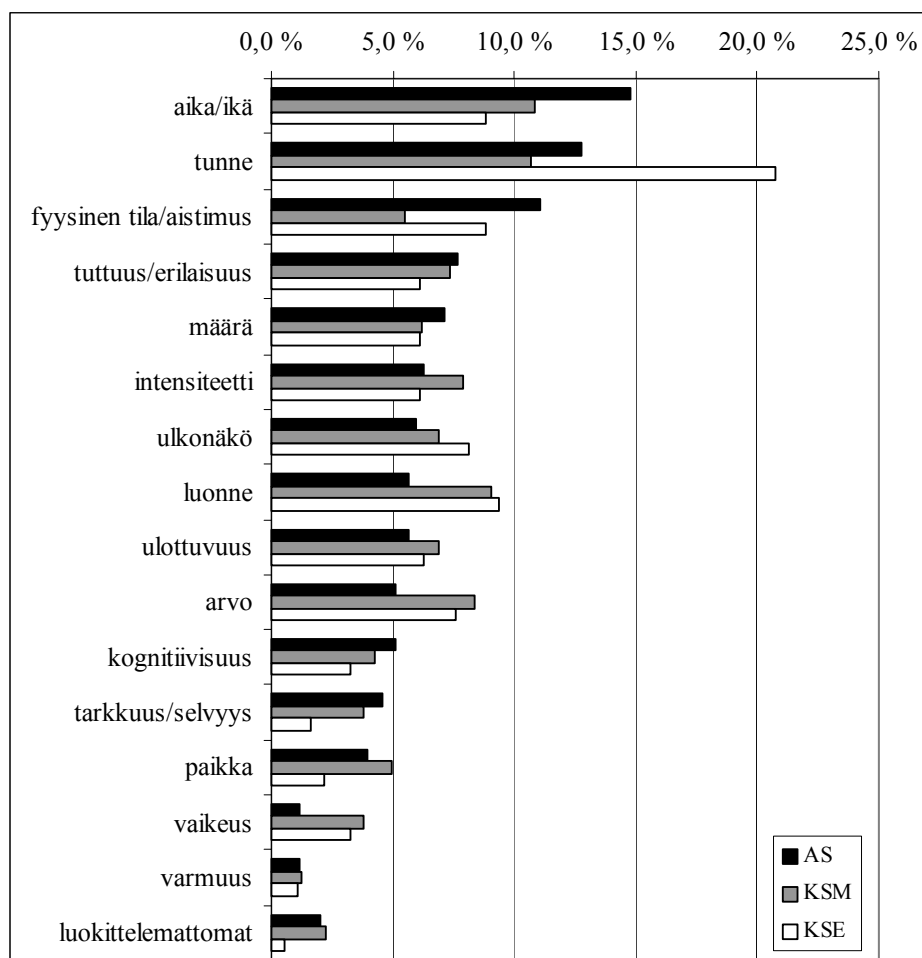
Semanttiseen preferenssiin liittyviä kontekstuaalisia valintoja tarkastellaan jälleen keskittyen pääsana- eli O1-positioon. Tässä positiossa olevia lekseemejä luokiteltaessa käännskielestä ei löytynyt sellaisia merkitysluokkia, joita ei myös tullut esiin, kun *hyvin*-määritteen semanttista preferenssiä tutkittiin aiemmin synonymia-analyysia varten. Tämä tarkoittaa sitä, että käännskielellä määritteen semanttisessa preferenssissä ei ole laadullisia eroja. Sen sijaan merkitysluokkien suhteellissa osuuksissa on tilastollisesti merkitseviä eroja, ja aineistosta on löydettävissä sekä kullekin osakorpukselle tyypillisiä että käännskielelle tyypillisiä frekvenssejä. *hyvin*-määritteen semanttisen preferenssin analyysi on esitetty tiivistetysti kuviossa 7.6.

Alkuperäis- ja käännsuomi erottuvat toisistaan kolmen merkitysluokan osuuden suhteen. Ensinnäkin vaikeutta ilmaisevien sanojen määrä on käännsuomen molemmissa aineistoissa merkitsevästi suurempi⁸⁰. Myös luonnetta ilmentävien osuus on käännsaineistoissa suurempi, ero on jälleen merkitsevä KSE- ja AS-aineistojen välillä ja toisessa vertailussa suuntaa-antava⁸¹. Aikaa ja ikää ilmentävien lekseemien osuus on molemmissa käännsuomen aineistoissa pienempi kuin AS-aineistossa; ryhmän osuus on merkitsevästi pienempi KSE-aineistossa ja KSM-aineistossa suuntaa-antavasti⁸². Tulos osoittaa, että aikaa ja ikää merkitsevien sanojen osuus on epätyypillisen pieni ja vaikeutta ja luonnetta ilmaisevien sanojen osuudet epätyypillisen suuria käännsuomessa. O1-kollokaateista, jotka on ryhmitelty edellä mainittuihin luokkiin, on esimerkkejä konkordanssissa 7.5 (s. 172).

⁸⁰ AS vs. KSM $z = 2,433097$ ($p < 0,05$); AS vs. KSE $z = 2,00992$ ($p < 0,05$).

⁸¹ AS vs. KSE $z = 2,001935$ ($p < 0,05$); AS vs. KSM $z = 1,900022$ ($p < 0,10$).

⁸² AS vs. KSE $z = 2,773094$ ($p < 0,01$); AS vs. KSM $z = 1,836065$ ($p < 0,10$).



Kuvio 7.6. *hyvin*: semanttinen preferenssi aineistoittain.

KSM-aineistosta erottuvat fyysistä tilaa ja aistimusta ilmaisevat sanat, joiden osuus on merkittävästi pienempi kuin muissa aineistoissa⁸³. Englannista suomennetut tekstit puolestaan erottuvat muista erittäin selvästi tunnetta mutta myös tarkkuutta ilmaisevien sanojen suhteen, erot ovat merkittäviä, käännösaineistojen välillä tunne-sanojen osalta peräti erittäin merkittäviä⁸⁴. Käännösaineistotkin poikkeavat siis jälleen toisistaan, eli englannista suomennettujen tekstien semanttinen preferenssi voi olla hieman erilainen kuin käännöskielessä yleensä. Tapauksista on esimerkkejä konkordanssissa 7.6.

⁸³ AS vs. KSM $z = 3,278476$ ($p < 0,01$); KSM vs. KSE $z = 2,307266$ ($p < 0,05$).

⁸⁴ tarkkuus: AS vs. KSE $z = 2,6210163$ ($p < 0,01$); KSM vs. KSE $z = 2,319284$ ($p < 0,05$).
tunne: AS vs. KSE $z = 3,055887$ ($p < 0,01$); KSM vs. KSE $z = 4,927054$ ($p < 0,001$).

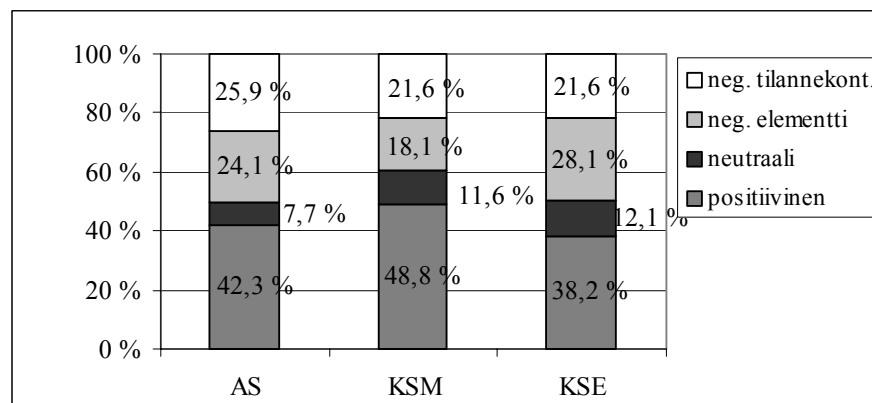
Konkordanssi 7.5. Esimerkkejä merkitysluokkiin aika/ikä (1–4), luonne (5–8) ja vaikeus (9–12) ryhmitellyistä kollokaateista.

AS		
1	uumis muistutti illan ja Tuijan saapuvan	<i>hyvinkin</i> pian . - Onko sinulla oma huuliharp
2	inen teollinen keksintö, vaan itse asiassa	<i>hyvin</i> vanha perinneruoka, jonka juuret s
3	. Minua oksetti ja lähdin ulos. Parasta oli	<i>hyvin</i> aikaisin aamulla. Haalistuva Etelän
4	astaan. - Sie oot saanu päähän tällin. - Jo	<i>hyvin</i> nuorena . - Mie soitan sairasauton.
KSE		
5	n mitä tehtäviini kuului. Mutta Nancy oli	<i>hyvin</i> ailahtelevainen , voisi sanoa kaksin
6	jokin toivesävelmä, ja hän oli minulle	<i>hyvin</i> huomaavainen ja kiltti, mistä McD
7	in kiusata Chuckieta. Hän tiesi olevansa	<i>hyvin</i> käytännönläheinen ihminen (totta
8	keä se millainen lohikäärme. Nämä kaksi	<i>hyvin</i> uskollinen , hyvä ji --- miten sanoo
KSM		
9	isen valituksensa. "Se on hyvin hankalaa	<i>hyvin</i> hankalaa totta tosiaan. Löysimme
10	maalaiskaupunkiinsa. Minun olisi ollut	<i>hyvin</i> helppo selittää hänen valintansa kei
11	väistämättä menee pilalle. Nyt minun on	<i>hyvin</i> vaikea tehdä koko juttua ymmärrett
12	varojärjestelmä on valmiina. Olisi ollut	<i>hyvin</i> yksinkertaista palauttaa kapteenin

Konkordanssi 7.6. Esimerkkejä merkitysluokkiin fyysinen tila/aistimus (1–4) ja tarkkuus/selvyys (5–8) ja tunne (9–12) ryhmitellyistä kollokaateista.

KSM		
1	tkään liikkumatta, hänen silmänsä olivat	<i>hyvin</i> auki ja sammuneen auringon valo
2	elta, joka ymmärtää kerran itse olleensa	<i>hyvinkin</i> haluttava ja tietää nykyään oleva
3	ut loistivat esikunnassa. Huoneessa oli	<i>hyvin</i> kuuma . Nainen vei Casey'n tuolin
4	polton hiukan kellastumat, mutta vielä	<i>hyvin</i> terveet hampaat paljastuivat: "Eve
KSE		
5	aluttomia viemään, mutta he vastailivat	<i>hyvin</i> seikkaperäisesti kyselyihimme t
6	inen tila. Ainakin sillä on hyvin vahva,	<i>hyvin</i> selvä metafyyminen ulottuvuus; jos
7	tä, että he kumpikin erottuivat joukosta	<i>hyvin</i> selvästi -- Ervin oli liian pitkä, Rç
8	non haju. Siihen aikaan sieraimet olivat	<i>hyvin</i> tarkat sellaisen suhteen. Sitä pait
KSE		
9	sanoi. Tämä oli järkyttävä tieto, todella	<i>hyvin</i> epämiellyttävä ajatus. "Eikö perh
10	un oli käytävä vessassa. Hän oli tästä	<i>hyvin</i> harmissaan , kunnes me molemm
11	sa vaarantamatta mainettaan. Nancy oli	<i>hyvin</i> kiihdyksissään , sillä herra Kinne
12	eita, mutta palanneita kasvoja? Olimme	<i>hyvin</i> onnellisia . Me kerroimme omat u

hyvin-määritteen viimeiseksi analysoitava kontekstuaalinen ominaisuus on semanttinen prosodia. Tätä varten on synonymia-analyysin tapaan tarkasteltu jälleen koko havaintoaluetta V3–O2 ja tarvittaessa tilannekontekstia. Prosodia on eritelty positiiviseen ja negatiiviseen, joista jälkimmäinen on edelleen jaettu negatiivisesta ja tilannekontekstista selviävään. Lisäksi omaksi ryhmäkseen on eritelty tapaukset, joiden prosodia on neutraali tai sitä ei ole pystytty selvittämään kontekstin avulla. Prosodiat on esitetty aineistoittain kuviossa 7.7.



Kuvio 7.7. *hyvin*: prosodia aineistoittain.

Negatiivisen ja positiivisen prosodian suhteet vaihtelevat aineistojen kesken siten, että AS- ja KSE-aineistoissa negatiivisen prosodian osuus on suurempi kuin positiivisen: AS-aineistossa negatiivisen osuus on tasan puolet 50,0 % ja KSE-aineistossa liki yhtä suuri, 49,7 %. KSM-aineistossa osuukien suhde on päinvastainen, sillä negatiivisen osuus on vain 39,7 %. Monilähdekielellisessä käänöskielessä *hyvin* näyttää siten olevan tyypillisemmin positiivisessa kontekstissa käytetty mutta muissa aineistoissa negatiivisessa, erot KSM-aineiston ja AS-aineiston välillä ovat tilastollisesti merkitseviä ja KSM- ja KSE aineistojen välillä erittäin merkitseviä⁸⁵.

Tulos viittaa siihen, että käänöskielessä prosodia voi olla kokonaisuutena osittain erilainen kuin alkuperäissuomessa, mutta tämä riippuu olla myös lähdekielestä. Negatiivisen prosodian vähyyden KSM-aineistossa näyttää johtuvan sellaisten tapauksien vähyydestä, jossa negatiivinen prosodia tulee esiin negatiivisen elementin avulla: tilannekontekstista selviävä negatiivinen prosodia ei erottele tilastollisesti merkitsevästi aineistoja, mutta negatiivisen elementtien osuus on KSM-aineistossa tilastollisesti merkitsevästi pienempi kuin muissa aineistoissa, suhteessa KSE-aineistoon jo erittäin merkitsevä⁸⁶. Käänöskielessä *hyvin*-määritteen kontekstista on siis löydettävissä vähemmän negatiivisia elementtejä, mutta tämä riippuu jälleen lähdekielestä. Konkordanssiin 7.7 on lueteltu esimerkkejä positiivisesta ja negatiivisesta prosodiasta; riveillä 1–5 esimerkkejä positiivisesta prosodiasta, negatiivisen aineksen (lihavoitu) aiheuttamia negatiivisen prosodian tapauksia riveillä 6–10 ja tilannekontekstin luomasta negatiivisesta prosodiasta esimerkkejä riveillä 11–15.

⁸⁵ Positiivinen: AS vs. KSM $z = 1,989536$ ($p < 0,05$); KSM vs. KSE $z = 3,767468$ ($p < 0,001$).

Negatiivinen: AS vs. KSM $z = 3,210825$ ($p < 0,01$); KSM vs. KSE $z = 3,778168$ ($p < 0,001$).

⁸⁶ Tilannekonteksti: AS vs. KSM $z = 1,557172$ ($p > 0,05$); KSM vs. KSE $z = 0,172036$ ($p > 0,05$).

Neg. elementti: AS vs. KSM $z = 2,333330$ ($p < 0,01$); KSM vs. KSE $z = 4,321198$ ($p < 0,001$).

Konkordanssi 7.7. Esimerkkejä *hyvin*-määritteen positiivisesta ja negatiivista prosodiasta.

AS			
1	Sii­ tä tuli oman pienen asiakaskuntansa keskuudessa	<i>hyvin</i>	suosittu. Tähän vaikutti mämmin perinteinen paastoruoan rooli.
2	Seksuaaliset ilmiuneni ovat	<i>hyvin</i>	realistisia ja pitkälle päivään saattavan mielihyvän sävyttämiä.
3	Niin Asko jäi virkavapaalle.	<i>Hyvinkin</i>	nopeasti hän huomasi kuinka paljon mielenkiintoista aikaa oli olemassa.
4	Hän tulee nykyään huoneeseeni	<i>hyvin</i>	harvoin, ja niin on hyvä.
5	Siinä kertoessa se päivä tuli	<i>hyvin</i>	lähelle, auringon sormet ihollani, ruohon tuoksu - puutarhuri leikkasi juuri nurmikkoo jossakin näköetäis
KSE			
6	Tämä oli järkyttävä tieto, todella	<i>hyvin</i>	epämiellyttävä ajatus.
7	Parissa sekunnissa he alkoivat voida	<i>hyvin</i>	huonosti . Poliisin varotoimet hidastivat pelastustoimien käynnistymistä.
8	Edellisen illan lasit ja lautaset olivat yhä pöydällä, ja ne näyttivät	<i>hyvin</i>	yksinäisiltä ja hylätyiltä , aivan kuin jokin hirmittävä tuho olisi kohdannut niitä jotka olivat niistä syöneet
9	Grace oli	<i>hyvin</i>	mustasukkainen koska isäntä suosi häntä enemmän taloudenhoitajaa, jota Grace taas vihasi ,
10	"Millainen sika! " sanoi isoäiti. "	<i>Hyvin</i>	paha poika", sanoi isoisä.
KSM			
11	Ja huoneita oli nyt viisi. Ja yksi niistä	<i>hyvin</i>	hämärä, koska ikkuna avautui miltei vasten naapuritalon seinää.
12	- Olet varmaankin oikeassa, mutta olet asettanut tavoitteen	<i>hyvin</i>	korkealle. Itse olen vain koettanut tottua itseeni. Sen pitemmälle
13	Keski-Aasian hiekkaerämaassa sattunut tapaus oli jättänyt salaperäisellä ja	<i>hyvin</i>	intimillä tavalla ikuiset jälkensä meidän sukumme historiaan. Huom
14	kirkasta valoa. Turbaanipäinen mies, pukeutunut pitkään takkiin, joka oli	<i>hyvin</i>	paksu ja hyvin kuuma siinä autiomaan hiekalla. Silmät vinot ja
15	Serenityä tuskin huomattiin hänen mustassa puvussaan, jossa oli kapeat käänteet, valkoisessa paidassaan, tummassa kielenpaksuisessa solmiossaan ja mustissa piikkikärkikengissään.	<i>Hyvin</i>	lyhyiksi leikattuine hiuksineen hänen päänsä näytti huomattavasti pienemmältä, olemuksensa suuremmalta ja laihemmalta, ja hänen korvansa muistuttivat oravan korvia

7.2.1.5. *hyvin*-määritteen kontekstuaalinen merkitysyksikkö kään­nös­suomessa

Edellä olevasta käsittelystä on kerätty taulukkoon 7.10 koonti niistä piirteistä, missä kohdin *hyvin*-määritteen kään­nös­suomalainen merkitysyksikkö eroaa alkuperäis-suomalaisesta. Taulukkoon on koottu epätyypilliset leksikaaliset myötäesiintymät ja

klusterit sekä muiden kontekstuaalisten piirteiden yli- ja aliedustuminen käännös-korpuksittain.

Taulukko 7.10. *hyvin*-määritteen epätyypilliset kontekstuaaliset kombinaatiot sekä piirteiden yli- ja aliedustuminen käännössuomessa (* suhteessa molempiin muihin aineistoihin).

	KSM-aineisto	KSE-aineisto
Kollokationaalinen valinta	Epätyypilliset kollokaatit*: ERIKOINEN, <i>harvoin</i> , HIENO, <i>ja</i> , KAUIAS, KORKEALLA, LYHYT, LÄHELLÄ (adp.) LÄHELLÄ (adv.), <i>nopeasti, paljon, pitkään</i> , SAIRAS, <i>selvästi</i> , SUURI, TYYTYVÄINEN, VAALEA, VAARALLINEN, VAATIMATON, VAHVA, VAIKUTTA, VAKAVA, <i>vakavasti</i> , <i>varhain</i> , ÄÄNI Epätyypilliset klusterit: (4) <MINÄ OLLA <i>hyvin</i> VAIKEA> (3) <HÄN OLLA <i>hyvin</i> > Klusteroitumissuunta vasemmalle	Epätyypilliset kollokaatit*: KALPEA, KUMMALLINEN, KYLMÄ, LÄHEINEN, NÄYTÄTÄ, ONNELLINEN, PAHA, PIMEÄ PITKÄ, SIISTI, <i>silti</i> , SURULLINEN, VIHAINEN, VOIMAKAS Epätyypilliset klusterit: (4) <HÄN OLLA <i>hyvin</i> PITKÄ> (3) <SE OLLA <i>hyvin</i> > Klusteroitumissuunta keskittynyt Yliedustuminen: - klusteroituminen 4-sanaisiin - klusteroituminen 3-sanaisiin
Kolligationaalinen valinta	Aliedustuminen: - kolligaattien klusteroituminen	Aliedustuminen: - substantiivikolligaatti - adverbikolligaatti - kvanttorikolligaatti - adpositiokolligaatti Yliedustuminen: - lauseenalkuisuuskolligaatti - lauseenloppuisuuskolligaatti Lyhyt lausepituus
Morfologinen preferenssi	Aliedustuminen: - liitepartikkeli <i>kin</i>	Aliedustuminen: - liitepartikkeli <i>kin</i>
Semanttinen preferenssi	Aliedustuminen: - fyysinen tila/aistimus Yliedustuminen: - vaikeus	Aliedustuminen: - aika/ikä - tarkkuus/selvyys Yliedustuminen: - vaikeus - luonne - tunne
Semanttinen prosodia	Aliedustuminen: - negatiivisen prosodia (neg. elementti) Yliedustuminen: - positiivisen prosodia	—

7.2.2. *kovin*

7.2.2.1. Kollokationaalinen valinta

kovin-määritteen tilastollisesti merkitsevät kollokaatit on lueteltu aineistoittain taulukossa 7.11.

Taulukko 7.11. Tilastollisesti merkitsevät *kovin*-määritteen kollokaatit osakorpuksittain (vain yhdessä aineistossa merkitsevät kollokaatit kursivoitu; *MI* = *MI*-testi, *t* = *t*-testi).

AS-aineisto			KSM-aineisto			KSE-aineisto		
kollokaatti	<i>MI</i>	<i>t</i>	kollokaatti	<i>MI</i>	<i>t</i>	kollokaatti	<i>MI</i>	<i>t</i>
usein	4,96	2,56	<i>kaukana</i> (adp.)	5,43	2,18	<i>pitkällä</i> (adv.)	4,87	2,16
tärkeä	4,53	2,14	<i>helposti</i>	4,73	2,15	<i>surullinen</i>	4,80	2,36
paljon	3,98	3,38	kuulostaa	4,47	2,34	pitkään	4,32	2,12
moni	3,76	3,07	paljon	4,13	6,04	paha	4,30	3,00
ei	2,22	7,11	kauan	3,90	3,49	kuulostaa	4,19	2,99
mutta	1,54	2,63	<i>tuskin</i>	3,72	2,45	paljon	3,76	4,90
olla	0,94	4,54	<i>vaikuttaa</i>	3,68	2,61	hyvin	3,62	3,18
			<i>iloinen</i>	3,68	2,06	<i>kovin</i>	3,57	2,42
			pitkään	3,66	2,44	kauan	3,53	2,42
			tärkeä	3,43	2,03	<i>nuori</i>	3,15	2,17
			paha	3,35	2,39	<i>kaukana</i> (adv.)	3,09	1,97
			<i>vaikea</i>	3,30	2,01	<i>pitkä</i>	2,73	2,40
			usein	3,15	2,35	hyvä	2,66	3,26
			<i>vähän</i>	3,01	3,03	moni	2,53	2,19
			moni	2,68	2,80	<i>ehkä</i>	2,44	2,58
			hyvin	2,61	2,21	ei	2,38	10,62
			<i>suuri</i>	2,30	2,39	tuntua	2,10	2,30
			tuntua	2,21	2,22	<i>sillä</i>	2,08	2,02
			<i>puhua</i>	2,06	2,15	mutta	1,48	3,69
			<i>näyttää</i>	1,96	2,35	olla	0,61	4,12
			ei	1,91	9,42			
			hyvä	1,81	2,26			
			mutta	1,39	4,19			
			<i>niin</i>	0,88	2,15			
			olla	0,82	6,23			

hyvin-määritteestä poiketen alkuperäissuomen aineistossa ei ole yhtään sellaista tilastollisesti merkitsevää kollokaattia, joka ei olisi merkitsevä myös jommassa-kummassa käännoaineistossa. Kaikille aineistolle yhteisiä kollokaatteja löytyy tämänkin määritteen kontekstista. Niitä on viisi:

kovin <EI, MONI, *mutta*, OLLA, *paljon*>.

Nämä ovat nimenomaan *kovin*-määritteelle tyypillisiä tilastollisesti merkitseviä kollokaatteja, jotka erottavat määritteen lähisyronyymeista (ks. aiempaa taulukkoa 6.4 s. 89). Niinpä myös käänöskielestä on löydettävissä juuri tälle määritteelle tyypillisimmät myötäesiintymät. Alkuperäissuomen aineistosta käänössuomen erottavat molemmassa käänösaineistoissa kollokaatit

kovin <*hyvin*, HYVÄ, *kauan*, KUULOSTAA, PAHA, *pitkään*, TUNTUA>.

Tosin näistä jokainen tavataan myös AS-aineistossa vähintään kerran, joten täysin tuntemattomia kollokaatioita ne eivät ole alkuperäissuomellekaan. Tyypillisinä niitä ei voi alkuperäissuomessa kuitenkaan pitää, joten ne viestivät alkuperäissuomelle epätyypillisistä kollokaatioista. Käänösaineistotkin eroavat toisistaan: KSM-aineistosta tavataan sekä O1-positiolla tyypillisiä kollokaatteja (*kovin* <*helposti*, ILOINEN, *kaukana*, SUURI, VAIKEA, *vähän*>) sekä myös muissa asemassa olevia kollokaatteja, jotka eivät ole merkitseviä englannista suomennetussa aineistossa. Viimeksi mainittuja ovat

kovin <*niin*, NÄYTTÄÄ, PUHUA, *tuskin*, VAIKUTTAA>.

Näistä VAIKUTTAA on tuttu jo *hyvin*-määritteen yhteydestä, jolloin se osoittautui niin ikään vain KSM-aineistolle tyypilliseksi. Sen sijaan NÄYTTÄÄ, joka oli *hyvin*-määritteen konteksteissa tyypillinen vain englannista suomennetussa aineistossa, onkin nyt tyypillinen KSM-aineistossa. Huomattakoon, että NÄYTTÄÄ-tapaukset eivät ole peräisin pelkästään KSM-aineiston englannista käännetystä materiaalista, vaan niitä tavataan viidestä eri lähdekielestä käännettyissä teksteissä. Kolmas verbi-kollokaatti PUHUA esiintyy *kovin*-määritteen kanssa yleensä tapauksissa, joissa *kovin* määrittää puhumisen tapaa tai määrää (konkordanssi 7.8 1–4). *niin*-kollokaateista suurin osa on tapauksia, joissa *niin* vahvistaa *kovin*-määritteen intensiteettiä ja tuo määrittäykseen myös modaalisen vivahteen (3–6) (*niin*-määritteen modaalisuudesta. Ahlman 1933; ks. myös Orpana 1973, Mäkipää 1995). Kollokaatti *tuskin* näyttää viihtyvän erityisesti käänössuomessa yhteisesiintymissä *kovin*-määritteen kanssa. Kollokaatio on täysin vieras sekä alkuperäissuomelle ja englannista käänetyille teksteille. Esimerkkejä *tuskin*-kollokaateista riveillä 7–9.

Konkordanssi 7.8. Esimerkkejä KSM-aineiston tyypillisistä *kovin*-määritteen kollokaateista.

1	ssa vain minä voin hyökätä? Puhumme	<i>kovin</i>	vähän huonosta omastatunnosta
2	ivät tänä ensimmäisenä iltana puhuneet	<i>kovin</i>	paljon, rupattelivat vain niitä näitä
3	oikien valtio puhui kirjailijoilleen niin	<i>kovin</i>	isällisesti, painoi heidät leveää rint
4	Kuulkaa", vanha mies puhui , "minä niin	<i>kovin</i>	mielelläni juttelin teidän kanssa
5	lumme." "Täällä kirjat eivät ole niin	<i>kovin</i>	suosittuja, poikaseni", hän sa
6	tö kulta, kun ajattelen Sinun ah niin	<i>kovin</i>	vaikeata suhdettasi (taidat olla
7	vahva persoona...? Sinulle on tuskin	<i>kovin</i>	mukavaa, että vanhus asuu teillä."
8	la jo iäkkäitä miehiä, jotka tuskin olivat	<i>kovin</i>	kiinnostavia noin nuoresta naisest
9	kirottua kultaanne. Tuskin olisi tarvittu	<i>kovinka</i>	kummoista valtiomiestä tai suurt

Kuten edelläkin, KSE-aineistostakin on löydettävissä omanlaisiaan sekä O1-positiivissa että muualla esiintyviä kollokaatteja, jotka erottavat aineiston kollokatio-naalisen valinnan muista aineistoista. Pääsanapositiivista löytyvät

kovin <NUORI, PITKÄ, *pitkälle*, SURULLINEN>.

Kollokaateista PITKÄ ja SURULLINEN tavattiin myös *hyvin*-määritteen kontekstissa tässä aineistossa. Siten kollokaatiot *hyvin*, *kovin* <PITKÄ, SURULLINEN> tuntuvatkin olevan juuri tälle aineistolle tyypillisiä. Muista positiivista tavattavia ovat kollokaatiot

kovin <*ehkä*, *kovin*, *sillä*>.

Näistä *kovin* tavataan toistorakenteissa ja rinnastuksissa (konkordanssissa 7.9 rivit 1–3), *ehkä* varausta ilmoittamassa kielteisissä yhteyksissä (4–6) ja *sillä* puolestaan *kovin* + O1-kollokaatti -rakenteen sisältävän lauseen sisältöä selittämässä (7–9).

Konkordanssi 7.9. Esimerkkejä KSE-aineiston tyypillisistä *kovin*-määritteen kollokaateista.

1	ssa, juonut liikaa diskoklubilla, paha asia	<i>kovin</i>	kovin	paha. Kun tulee takaisin, ei
2	nen työtovereiltaan kaupasta. "Voi, kovin	<i>kovin</i>	sievä",	hän kujertaa ja panee sen m
3	me kuivia, emme kovin nälkäisiä emmekä	<i>kovin</i>	viluisia. Oljet ovat hyvää eristysain	
4	öjen vessaan ei ollut ehkä sittenkään ollut	<i>kovin</i>	hyvä idea. Jatkoimme matkaa, kun	
5	seita, paljon ja halvalla. Ne eivät ehkä ole	<i>kovin</i>	luotettavia, mutta niihin käyvät kai	
6	oin sanoa kuka kuvaa." "Hän ei ehkä ole	<i>kovin</i>	valokuvauksellinen." "Ihanko totta.	
7	e ole alun alkaenkaan olleet siinä mielessä	<i>kovin</i>	ahneita (sillä me hyväksymme ede	
8	dä kauanko hän on poissa; toivottavasti ei	<i>kovin</i>	pitkään, sillä minä olen jotenkin tott	
9	ääriviivoista. Hän toivoo ettei nainen ole	<i>kovin</i>	vanha, sillä se jotenkin pahentaa s	

Frekventeimpien kollokaattien järjestys osoittautuu aineistoittain taulukon 7.12 mukaiseksi. Taulukkoon on lueteltu jälleen 20 taajimmin esiintyvää kollokaattia ja kaikki vähintään viisi kertaa esiintyvät on lueteltu liitteessä 3 (taulukko 2). Huomio kiinnittyy välittömästi mielenkiintoiseen seikkaan: englannista käännettyissä teksteissä yleisin kollokaatti on EI, kun muissa se on OLLA. Näiden kahden kollokaatin yleisysero on KSE-aineistossa lisäksi hyvin selvä. EI-kollokaatin korostunut merkitys englannista käännettyissä teksteissä saa siis lisätodistusta, edellähän EI osoittautui poikkeuksellisen merkittäväksi kollokaatiksi juuri tässä aineistossa. Ero johdetaan ainakin kahdesta syystä: Ensinnäkin elliptisistä finiittiverbittömistä rakenteista, joissa *kovin* esiintyy KSE-aineistossa kaksi kertaa useammin (15,4 %) kuin esimerkiksi KSM-aineistossa (7,9 %). Näistä rakenteista on hyvin usein jätetty pois OLLA-verbi, mikä siis vähentää kyseisen kollokaatin osuutta. Toinen syy löytyy toistorakenteista, joissa EI mahtuu englannista käännettyissä teksteissä tarkastelualueelle kahdesti, toisinaan jopa kolmasti. Esimerkkejä molemmista tapauksista konkordanssissa 7.10.

Taulukko 7.12. Astemäärityksen *kovin* vähintään 5 kertaa tarkastelualueella V3–O2 esiintyvistä kollokaateista 20 yleisintä. f = frekvenssi, f % = suhteellinen frekvenssi kaikista kollokaateista, F = summafrekvenssi, F % = suhteellinen summafrekvenssi. (Suhteelliset osuudet on laskettu kollokaattien yhteismäärästä, jotka ovat AS-aineistossa 880, KSM-aineistossa 2070 ja KSE-aineistossa 1530.)

AS-aineisto				KSM-aineisto				KSE-aineisto						
	f	f%	F	F%		f	f%	F	F%		f	f%	F	F%
olla	90	10,2	90	10,2	olla	205	9,9	205	9,9	ei	173	11,3	173	11,3
ei	82	9,3	172	19,5	ei	165	8,0	370	17,9	olla	144	9,4	317	20,7
se	31	3,5	203	23,1	se	53	2,6	423	20,4	se	51	3,3	368	24,1
hän	18	2,0	221	25,1	ja	47	2,3	470	22,7	hän	39	2,5	407	26,6
mutta	16	1,8	237	26,9	mutta	46	2,2	516	24,9	mutta	33	2,2	440	28,8
ja	15	1,7	252	28,6	hän	42	2,0	558	27,0	paljon	28	1,8	468	30,6
minä	13	1,5	265	30,1	paljon	41	2,0	599	28,9	ja	20	1,3	488	31,9
paljon	13	1,5	278	31,6	että	26	1,3	625	30,2	minä	17	1,1	505	33,0
että	11	1,3	289	32,8	minä	24	1,2	649	31,4	että	16	1,0	521	34,1
moni	11	1,3	300	34,1	niin	22	1,1	671	32,4	niin	16	1,0	537	35,1
niin	8	0,9	308	35,0	he	15	0,7	686	33,1	hyvä	15	1,0	552	36,1
usein	7	0,8	315	35,8	kauan	14	0,7	700	33,8	hyvin	12	0,8	564	36,9
he	5	0,6	320	36,4	me	12	0,6	712	34,4	ehkä	10	0,7	574	37,5
mies	5	0,6	325	36,9	vähän	12	0,6	724	35,0	kuulostaa	10	0,7	584	38,2
nainen	5	0,6	330	37,5	joka	11	0,5	735	35,5	paha	10	0,7	594	38,8
tämä	5	0,6	335	38,1	moni	11	0,5	746	36,0	sanoa	9	0,6	603	39,4
tärkeä	5	0,6	340	38,6	hyvä	10	0,5	756	36,5	tuntua	9	0,6	612	40,0
voida	5	0,6	345	39,2	näyttää	10	0,5	766	37,0	tämä	9	0,6	621	40,6
vuosi	5	0,6	350	39,8	tämä	10	0,5	776	37,5	pitkä	8	0,5	629	41,1
					vain	10	0,5	786	38,0	sinä	8	0,5	637	41,6

Konkordanssi 7.10. Esimerkkejä KSE-aineiston EI-kollokaatin osuutta lisäävistä kombinaatioista.

1	än elämä ei ole kovinkaan uusi eikä <i>kovin</i> vieras. Varsinkin kun on vakoillut <i>mikään maa</i> tuon maa
3	Silläkin puolella on valtameri, äiti sanoo niin, eikä <i>edes</i> <i>kaukana</i> , iso kylmä valtameri, ja Buddy näkee tuntemattoman
4	ä Nancy oli ennen palvelijana tässä <i>kovinkaan kauan sitten</i> , ja teki kaikki raskaat j talossa, ei
5	Kwan sanoo. " <i>Changmianissa ei ehkä</i> <i>kovin hienoa</i> ." "Se sopii minulle." Kwan
6	Se olisi vain hiukan uninen, ei muuta, ei <i>kovin</i> puhelias, eikä se muistaisi mitään s
7	la-Wossiky tunsu taas vihreän maan, ei <i>kovin</i> selvästi, ei niin kuin ennen, mutta
8	hden vuoden iässä mikään elämä ei ole <i>kovinkaan</i> uusi eikä mikään maa <i>kovin</i> vieras.

Käännösuomessa *kovin*-määrityksen kotekstista on siis löydettävissä sekä alkupe- räissuomelle epätyypillisiä kollokaatioita että myös pelkästään englannista suomen- netuille teksteille tyypillisiä kollokaatioita. Myös kollokaatioita pitemmät klusteri- rakenteet poikkeavat toisistaan jossain määrin eri aineistoissa (taulukko 7.13).

Taulukko 7.13. *kovin*-määritteen klusteroituminen aineistoittain.

	sanoja	V3	V2	V1	noodi	O1	O2	f
AS	4	hän	ei	olla	<i>kovin</i>			3
		minä	ei	olla	<i>kovin</i>			2
		ei	se	olla	<i>kovin</i>			2
		olla	viime	aika	<i>kovin</i>			2
	3		olla	minä	<i>kovin</i>	tärkeä		2
			ei	olla	<i>kovin</i>			24
			se	olla	<i>kovin</i>			7
				olla	<i>kovin</i>	moni		4
				olla	<i>kovin</i>	paljon		4
KSM	5	ei	olla	olla	<i>kovin</i>	moni		2
	4	se	ei	olla	<i>kovin</i>			10
		ei	se	olla	<i>kovin</i>			2
		ei	olla	edes	<i>kovin</i>			2
		ei	olla	olla	<i>kovin</i>			2
		ei	olla	olla	<i>kovin</i>	paha		3
			olla	olla	<i>kovin</i>	moni		2
		ei	olla	olla	<i>kovin</i>	kauan		2
		ei	olla	olla	<i>kovin</i>	kaukana		2
		ei	olla	olla	<i>kovin</i>	suuri		2
		ei	olla	olla	<i>kovin</i>	todennäköinen		2
		ei	olla	olla	<i>kovin</i>	vaikea		2
			tuskin	on	<i>kovin</i>	paljon		2
				olla	<i>kovin</i>	kaukana	se	3
				olla	<i>kovin</i>	suuri	mutta	2
				muuttua	<i>kovin</i>	paljon	hän	2
	3		ei	olla	<i>kovin</i>			36
				olla	<i>kovin</i>	paljon		11
				olla	<i>kovin</i>			7
				olla	<i>kovin</i>			7
KSE	5	se	ei	olla	<i>kovin</i>	kummoinen		2
	4	se	ei	olla	<i>kovin</i>			7
		että	se	olla	<i>kovin</i>			3
		ei	ehkä	olla	<i>kovin</i>			2
		että	minä	olla	<i>kovin</i>			2
		vesi	ei	olla	<i>kovin</i>			2
			ei	olla	<i>kovin</i>	hyvä		3
			ei	olla	<i>kovin</i>	kummoinen		3
			ei	kestää	<i>kovin</i>	kauan		2
			ei	olla	<i>kovin</i>	huono		2
			ei	olla	<i>kovin</i>	korkealla		2
			se	olla	<i>kovin</i>	surullinen		2
	3		ei	olla	<i>kovin</i>			36
				olla	<i>kovin</i>	hyvä		8
			mutta	ei	<i>kovin</i>			7

Viisisanaisia klustereita on pelkästään käänösaineistoissa, mikä viittaa käännöskielen kiinteämpiin leksikaalisiin suhteisiin ja mahdollisesti käännöskielen yk-

sinkertaistumiseen. Nelisanaisten klustereiden mukaan laskettu klusteroitumisaste on AS:ssa 6,3, KSM:ssa 9,7 ja KSE:ssa 9,8 ja kolmiosaisien vastaavassa järjestyksessä 38,6, 35,5 ja 40,8. Erot eivät ole merkitseviä kummassakaan tapauksessa käännös- ja alkuperäissuomen välillä⁸⁷, eli *kovin*-määritteen klusteroituminen ei tue oletusta käännöskielen yksinkertaistumisesta.

Sen sijaan klusterirakenteissa on havaittavissa variaatiota, mikä kertoo alkuperäissuomelle epätyypillisistä leksikaalisista valinnoista. Positiossa V3 frekventeimmän nelisanaisen klusterin lekseemi on nimittäin AS-aineistossa erilainen kuin käännösaineistossa: positiosta löytyy pronomini HÄN ja myös MINÄ kahdesti, mutta käännösaineistoista kumpaakaan pronominia ei tavata tästä positiosta vaan tyypillisin on SE. Taajimmin tavattava nelisanainen klusteri on molemmissa käännösaineistoissa <SE EI OLLA *kovin*>, mutta AS-aineistossa <HÄN EI OLLA *kovin*>. V1- ja V2-positioiden tyypillisimpiä lekseemejä EI ja OLLA riippumatta aineistosta. O1-positioiden taas selvästi vaihtelua aineistojen kesken: alkuperäissuomessa ei ensinnäkään juuri tavata muutamaa poikkeusta lukuun ottamatta klustereita, jotka yltävät tähän positioon, käännösaineistossa nämä ovat sen sijaan tavallisia; myös lekseemit vaihtelevat aineistoittain tässä positioiden. KSM poikkeaa muista vielä siinä, että ainoastaan tässä aineistossa klusterit yltävät O2-asemaan asti. Frekventein kolmisanaisten klusteri on kaikissa aineistoissa sama, sen sijaan näitä harvemmin tavattavissa klustereissa on leksikaalista vaihtelua jonkin verran.

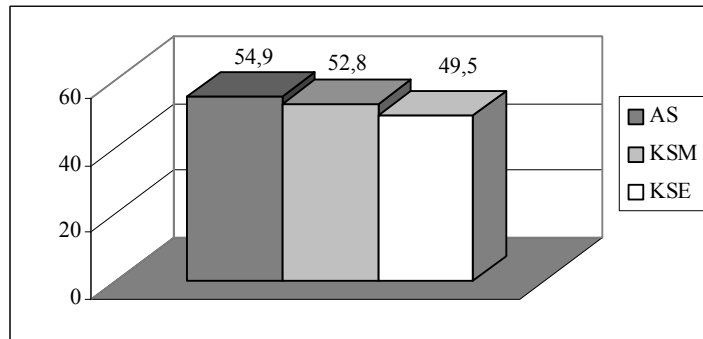
Kuten aiemminkin, yksinkertaistumishypoteesin testaamista varten tarkastellaan yleisimpien kollokaattien osuutta (taulukko 7.14) ja lemma–sane-suhteita koko tarkastelualueella (kuvio 7.8) sekä jälkimmäistä myös positiokohtaisesti (kuvio 7.9). Tilastollisen testauksen perusteella näyttää ensinnäkin siltä, että yleisimpien kollokaattien osuudet eivät ole suurempia käännöksissä kuin alkuperäissuomalaisissaakaan teksteissä. Ainoana poikkeuksena tähän on KSE-aineiston viiden yleisimmän kollokaatin suurempi osuus KSM-aineistoon verrattuna, mikä viittaa jonkinasteiseen lähdekielen vaikutukseen. Havainnot frekventeimpien kollokaattien suuremmasta peitosta ja siten yksinkertaistumisesta käännöskieleessä (Laviosa-Braithwaite 1996; Nevalainen tulossa) eivät saa tuloksesta juurikaan tukea, joskin on edelleen pidettävä mielessä, että käsillä olevassa tutkimuksessa ei tarkastella koko aineistoa koskevaa leksikaalista yksinkertaistumista.

Taulukko 7.14. *kovin*: yleisimpien kollokaattien summafrekvenssien suhteellisten osuuksien vertailu.

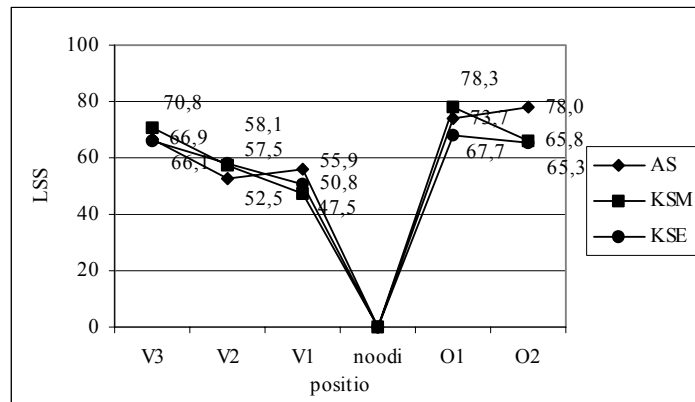
	AS vs. KSM	AS vs. KSE	KSM vs. KSE
5 yleisimmän osuus	26,9 % vs. 24,9 % z = 1,14, p > 0,05	26,9 % vs. 28,8 % z = 0,96, p > 0,05	24,9 % vs. 28,8 % z = 2,57, p < 0,05
10 yleisimmän osuus	34,1 % vs. 32,4 % z = 0,89, p > 0,05	34,1 % vs. 34,1 % z = 0,02, p > 0,05	32,4 % vs. 34,1 % z = 1,03, p > 0,05

⁸⁷ Nelisanaiset: KSM vs. AS z = 1,349246 (p > 0,05); KSE vs. AS z = 1,346596 (p > 0,05).
Kolmisanaiset: KSM vs. AS z = 0,722553 (p > 0,05); KSE vs. AS z = 0,477480 (p > 0,05).

Kuviossa 7.8 esitetty koko tarkastelualueen lemman–sane-suhde antaa viitteitä kotekstin leksikaalisesta yksinkertaistumisesta, sillä LSS on molemmissa käänös-aineistoissa alkuperäissuomen arvoa pienempi. Tilastollinen vertailu kuitenkin osoittaa, että aineistojen väliset erot eivät ole tälläkään määritteellä tilastollisesti merkitseviä koko aluetta tarkasteltaessa⁸⁸. Kollokationaalinen diversiteetti näyttää varioivan silmämääräisesti arvioiden eniten oikeanpuoleisissa positioissa (kuvio 7.9). Tilastollisesti merkitseviä eroja löytyykin positioista O2, jossa alkuperäissuomen aineiston kollokationaalinen diversiteetti on suurempi kuin molemmissa käänöskorpuksissa⁸⁹. Aineistojen välisille eroille ei ole löydettävissä perustetta tietyn leksemin yliedustumisesta käänösaineistossa juuri tässä positiossa, vaan erot johtuvat position koko kollokationaalisen variaation erilaisuudesta.



Kuvio 7.8. *kovin*: lemman–sane-suhde tarkastelualueella V3–O2.



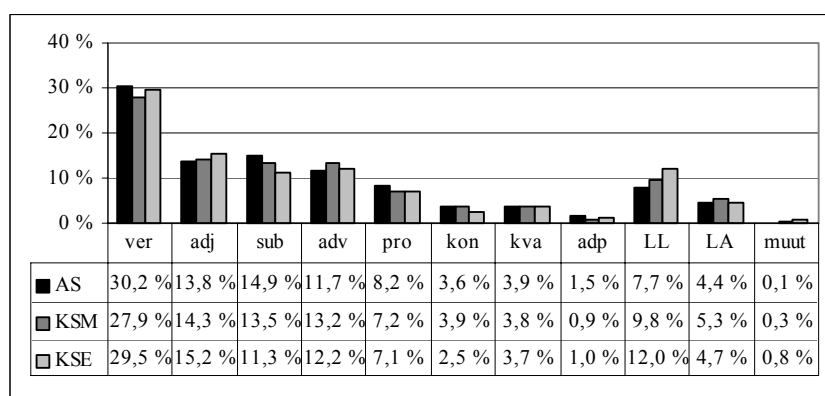
Kuvio 7.9. *kovin*: lemman–sane-suhde positioittain.

⁸⁸ AS vs. KSM $z = 0,739173$ ($p > 0,05$); AS vs. KSE $z = 1,879247$ ($p > 0,05$); KSM vs. KSE $z = 1,159004$ ($p > 0,05$).

⁸⁹ AS vs. KSM $z = 2,047739$ ($p < 0,05$); AS vs. KSE $z = 2,177609$ ($p < 0,05$).

7.2.2.2. Kolligationaalinen valinta

hyvin-määritteen yhteydessä ilmeni, että määritteen kontekstissa ei ollut epätyypillisiä kolligationaalisia suhteita, mutta joidenkin kolligaattien ja yleisimpien kolligaattiklusterien laadullisia ja määrällisiä eroja oli havaittavissa. Myös kontekstuaalinen lausepituus erotteli aineistot toisistaan. Erityisesti korostui englannista käännettyjen tekstien erilaisuus muihin aineistoihin verrattuna. *kovin*-määritteen sanaluokkakolligaatioita koskeva valinta on kuvattu aineistoittain kuviossa 7.10.



Kuvio 7.10. *kovin*: kolligaatit tarkastelualueella V3–O2.

Jälleen AS- ja KSM-aineistojen väliset kolligaattien osuuksien erot ovat hyvin pienet, eivätkä ne myöskään osoittaudu tilastollisesti merkitseviksi (taulukko 7.15). Englannista käännettyt tekstit sen sijaan eroavat alkuperäissuomalaisista teksteistä substantiivi- ja lauseenloppuuskolligaatin osuuksissa: edellisen osuus on merkitsevästi pienempi ja jälkimmäisen suurempi KSE-aineistossa. Erityisen suuri substantiivikolligaatin ero on positiossa V3: osuus on AS-aineistossa 29,0 % (n = 51) ja KSE-aineistossa lähes puolta vähemmän 16,0 % (n = 49), ero on myös tilastollisesti erittäin merkitsevä⁹⁰. Lauseenloppuus koskee luonnollisesti positiota O2. Se erottaa myös käännösaaineistot toisistaan, toisena erottavana kolligaattina ovat konjunktiot. Konjunktioiden osuuksissa on selkeitä eroja kahdessa positiossa: KSE:ssä ei ole V1 positiossa ainuttakaan konjunktiokolligaattia ja O2-positioissakin vain 1,3 % (n = 4), KSM:ssä vastaavat osuudet ovat 3,9 % (n = 16) ja 4,1 % (n = 17). Erot ovat tilastollisesti merkitseviä⁹¹. KSM-aineistossa konjunktiot ovat tavallisimmin rinnastuskonjunktioita joko rinnastamassa lausekkeita positioissa V1 ja O2 (konkordanssissa 7.11 rivit 1–4) ja aloittamassa lausetta positiossa V1 (5–6) tai alistuskonjunktiota lauseen alussa (7–8). KSE:ssä tällaisia esiintymiä ei siis juurikaan ole.

⁹⁰ AS vs. KSE $z = 3,379517$ ($p < 0,001$).

⁹¹ V1-positiossa AS vs. KSE $z = 3,477767$ ($p < 0,001$); O2-positiossa AS vs. KSE $z = 2,206459$ ($p < 0,05$).

Taulukko 7.15. *kovin*: kolligaatit, joiden osuudet erottavat aineistot toisistaan tilastollisin perustein koko tarkastelualueella.

AS - KSM	AS - KSE	KSM - KSE
	substantiivit	konjunktiot
–	$z = 2,55, p < 0,05$	$z = 2,25, p < 0,05$
	lauseenloppuisuus	lauseenloppuisuus
	$z = 3,28, p < 0,01$	$z = 2,11, p < 0,05$

Konkordanssi 7.11. Esimerkkejä konjunktiokolligasteista KSM-aineistossa.

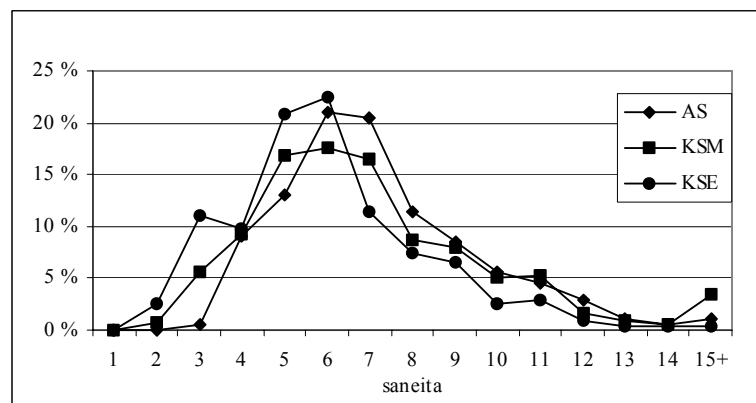
1	Martha kyllä siihen aikaan vielä puhui <i>kovin</i> varovasti 'tietyistä vaikeuksista so-
	aivan puolueen hengessä 'meidän tasaval-
	lastamme' ja
2	Leslie-Gordonin, jota sitten kauniin mutta <i>kovin</i> hermostuneen Cécilen takia tähtäsi
3	ista uhkakuvista, sillä nuoruudenystävä oli <i>kovin</i> luutunut ja ahdasmielinen; ja hänen
4	usiin, joka totta puhuen ei vaikuttanut enää <i>kovin</i> uudelta vaan hiukan kuluneelta, mut
5	in tätä ennen olin huomannutkaan. Mutta <i>kovin</i> pitkään minulla ei ollut aikaa katsel
6	ivat hänelle melko helposti. Mutta... Mutta <i>kovin</i> epätavallista asiaa tuo vasta Moskov
7	Sillä emme hirvinneet kuvitellakaan, että <i>kovin</i> kovin pian me jäämme ilman niitä k
8	mutta enää en muista sitä ollenkaan. Sillä <i>kovin</i> pitkää aikaa niiden toimeenpanemis

Kuten *hyvin*-tapauksessakin, tyypillisin kolligaattiklusteri on kaikissa aineistoissa aivan sama:

<SUB VER VER *kovin* ADJ LL>,

lisäksi tällä kertaa klustereiden osuudetkin ovat lähes samanlaiset: osuudet ovat AS:ssa 5,1 % (n = 9), KSM:ssa 3,4 % (n = 14) ja KSE:ssa 4,2 % (n = 13). Kolligaattien klusteroituminen ei siten tuo esille epätyypillisiä kombinaatioita eikä myöskään erätyypillisiä frekvenssejä ja yksinkertaistumista käännessuomessa.

Edellä selvisi, että lauseenloppuisuuden osuus on KSE-aineistossa suurempi kuin muissa aineistossa. Tilanne oli samanlainen myös *hyvin*-tapauksessa. Lauseenloppuisuuskolligaatin suurempi osuus saattaa myös tässä tapauksessa olla merkki lyhyemmästä lausepituudesta. Tyypillisin lauseiden sanemäärä on kuitenkin jokaisessa aineistossa sama, kuusi sanetta (AS n = 37, KSM n = 73, KSE n = 69). Sitä lyhyempien lauseiden osuus näyttäisi silti olevan KSE-aineistossa suurempi kuin muissa aineistossa ja pitempien lauseiden osuus taas hieman pienempi (kuvio 7.11). Lauseiden sanemäärien keskiarvot (taulukko 7.16) paljastavatkin englannista suomennettujen tekstien *kovin*-määritteen kotekstilauseet selvästi lyhyemmiksi kuin muissa aineistoissa. Erot ovat kahteen muuhun aineistoon verrattuna tilastollisesti erittäin merkitseviä. Tulos on samanlainen kuin *hyvin*-määritteelläkin: englannista suomennetuissa teksteissä lauseet, joissa määritteitä on käytetty, ovat lyhyempiä kuin yleensä käännesskielessä ja myös lyhyempiä kuin alkuperäissuomessa.

Kuvio 7.11. *kovin*: lauseiden sanemäärä.Taulukko 7.16. *kovin*: lauseiden sanemäärien keskiarvot ja hajonta sekä keskiarvojen erojen tilastollinen merkitsevyys aineistoittain.

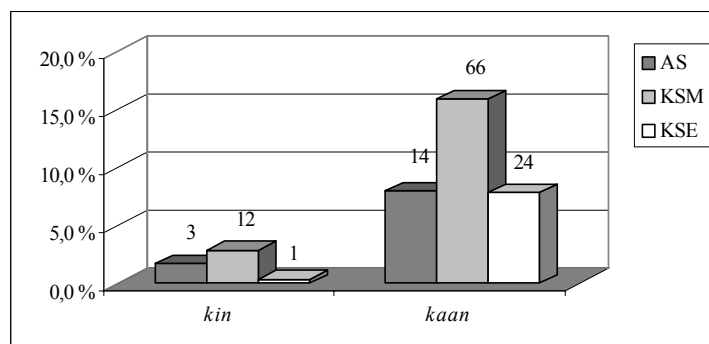
	AS	KSM	KSE
keskiarvo (saneita)	7,28	7,11	6,01
vaihteluväli (saneita)	19 [3–22]	21 [2–23]	13 [2–15]
	AS–KSM	AS–KSE	KSM–KSE
merkitsevyys	$z = 0,71$ ($p > 0,05$)	$z = 5,34$ ($p < 0,001$)	$z = 5,49$ ($p < 0,001$)

7.2.2.3. Morfologinen preferenssi

Päinvastoin kuin *hyvin*-tapauksessa, jossa liitepartikkelien määrä oli pienempi käännoissuomessa, *kovin*-tapauksessa niiden määrä on käännoississä yleensä suurempi kuin alkuperäissuomessa, mutta englannista suomennetuissa teksteissä osuus on likipitään sama kuin AS-aineistossa (kuvio 7.12). Molemmat mahdolliset liitepartikkelit antavat silmämääräisesti samansuuntaisen tuloksen, mutta *kin*-liitepartikkelin osuuksista ainoastaan ero käännoisaineistojen välillä on tilastollisesti merkitsevä. *kaan*-liitepartikkelin määrät sen sijaan eroavat tilastollisesti merkitsevästi sekä AS- ja KSM- että KSM- ja KSE-aineistojen välillä⁹². KSM-aineistossa *kaan*-liitepartikkelillisten *kovin*-määritteiden osuus on 15,9 % eli kaksinkertainen AS-aineiston (8,0 %) ja KSE-aineiston (7,8 %) osuuksiin nähden. Toisin kuin *hyvin*-tapauksessa, tulos ei tue oletusta kohdekielispesifisten aineiden aliedustumisesta käännoiskielessä, koska käännoismateriaaleissa osuudet ovat yhtä suuret tai suurem-

⁹² AS vs. KSM $z = 2,592761$ ($p < 0,10$); AS vs. KSE $z = 0,043699$ ($p > 0,05$); KSM vs. KSE $z = 3,248338$ ($p < 0,10$).

mat kuin alkuperäissuomessa. KSM-aineistossa esiintymiä on yhtä hyvin virosta ja unkarista kuin indoeurooppalaisistakin kielistä käännettyissä teksteissä.

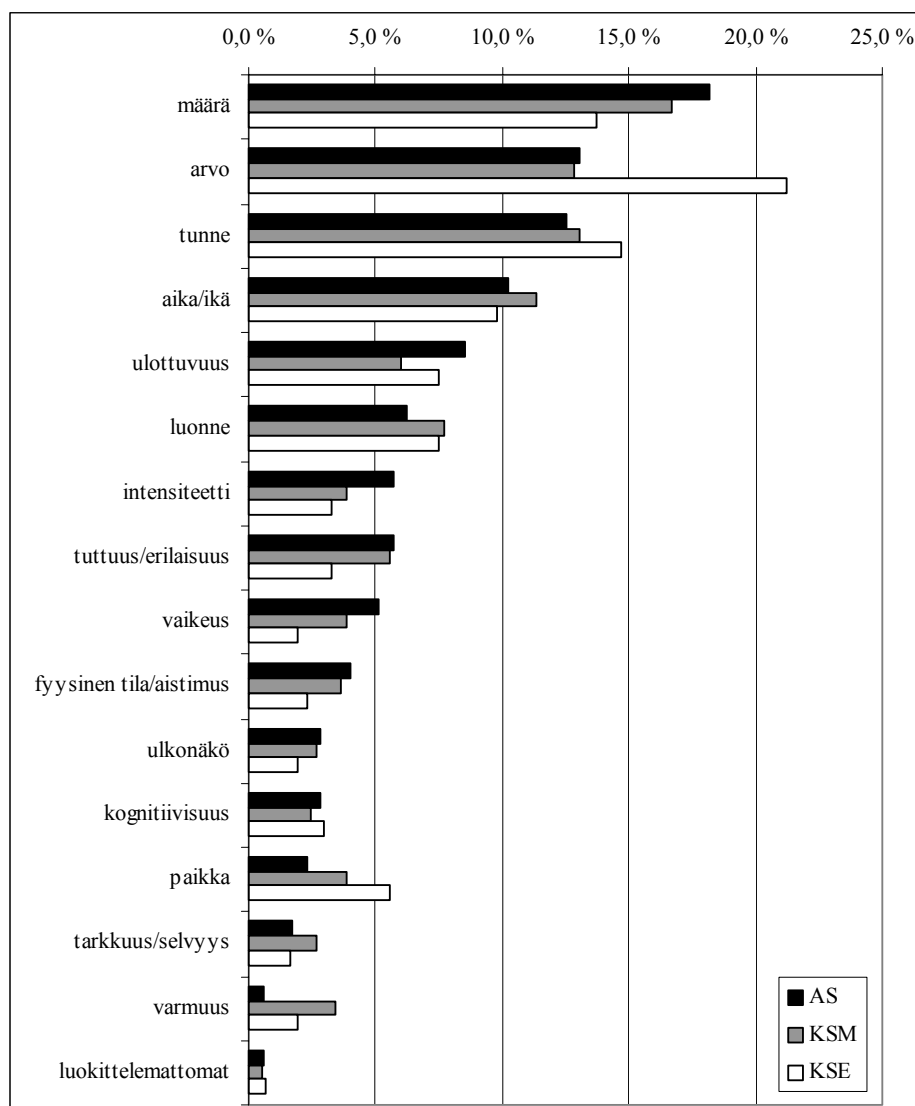


Kuvio 7.12. *kovin*: liitepartikkelien (*kin*, *kaan*) osuus aineistoittain.

7.2.2.4. Semanttinen preferenssi ja prosodia

kovin-määritteen semanttinen preferenssi on kuvattu aineistoittain kuviossa 7.13. Määritteen kontekstin O1-positiosta ilmenevät semanttiset preferenssit ovat erittäin samankaltaisia kaikissa aineistoissa. Tilanne on siis varsin erilainen kuin *hyvin*-määritteen semanttista preferenssiä tarkasteltaessa, jolloin aineistojen väliltä löytyi useita semanttisia luokkia, joiden osuudet vaihtelivat aineistojen välillä. *kovin*-määritteen alkuperäissuomalaisissa teksteissä ilmenevät preferenssit eivät ensinnäkään poikkea käännöskielen (KSM) preferensseistä, joten käännössuomi osoittautuu tässä kohtaa samanlaiseksi kuin alkuperäissuomikin. Myös käännösaineistot ovat keskenään hyvin samanlaisia, niiden pääsanapositiosta esiintyvät preferenssit poikkeavat ainoastaan arvo-merkitysluokan osuuksissa tilastollisesti merkitsevästi. Kyseinen merkityspiirre erottaa KSE-aineiston myös alkuperäissuomalaisista teksteistä. ”Arvosanat” näyttävät siten olevan englannista suomennetuissa teksteissä selvästi useammin *kovin*-määritteen pääsanana kuin muissa aineistoissa.⁹³ Osuuk-sien ero johtuu runsaista *hyvin*- ja *paha/pahasti*-sanojen esiintymistä, joista sekä *paha* että *hyvin* ovat *kovin*-määritteen tyypillisiä kollokaatteja (ks. aiemmin taulukkoa 7.11). Merkitysluokasta on esimerkkejä konkordanssissa 7.12.

⁹³ AS vs. KSE $z = 2,236491$ ($p < 0,05$); KSM vs. KSE $z = -3,024301$ ($p < 0,01$).



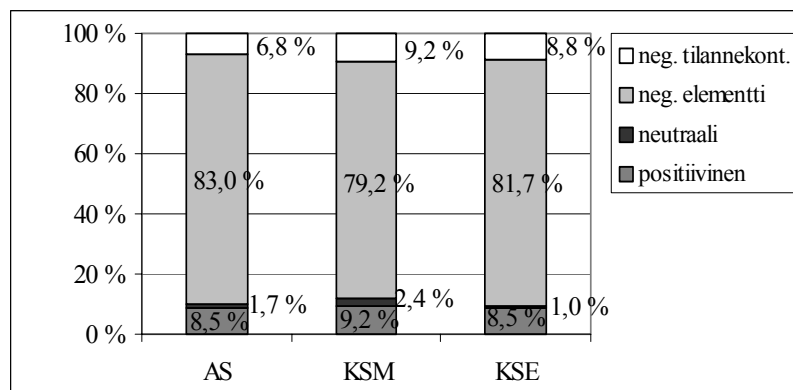
Kuvio 7.13. *kovin*: semanttinen preferenssi aineistoittain.

Edellinen analyysi siis osoittaa *kovin*-määritteen semanttisen preferenssin O1-positiossa hyvin samankaltaiseksi kaikissa aineistoissa. Samaa todistaa myös semanttisen prosodian tarkastelu. Sekä positiivisen että negatiivisen prosodian osuudet ovat lähes samansuuruisia kaikissa aineistoissa (kuvio 7.14). Negatiivisen prosodian, jonka osuus on selvästi suurempi, osuus vaihtelee KSM-aineiston 88,4 prosentista KSE-aineiston 90,5 prosenttiin, AS-aineistossa osuus on 89,8 %. Negatiivisen prosodian sisäinen jakautuminen on myös hyvin samanlaista aineistoittain. *ko-*

vin-määritteen semanttiset kontekstipiirteet näyttävät olevan siten hyvin jähmeitä muuttumaan kääntämisessä, eikä tähän juuri vaikuta lähdekielikään.

Konkordanssi 7.12. Esimerkkejä arvo-merkitysluokkaan ryhmitellyistä kollokaateista.

KSE			
1	ta. "Otit sen puheeksi minun kannaltani	<i>kovin</i>	huonolla hetkellä." Andy katsoi O
2	tä hän on maalta kotoisin eikä sota ollut	<i>kovin</i>	paha kohteleminen, vaikka saksalai
3	juutalaiset, tekivät kuolemaa, enkä ollut	<i>kovinkaan</i>	hyvin selvillä heidän kohtalostaan
4	siitä varma, vaikkei edes tuntenut häntä	<i>kovin</i>	hyvin . Ehkä Andy Thatcher ei vai
5	paaneja, mutta tulos ei ollut esteettisesti	<i>kovinkaan</i>	tydyttävä . Korvaukseksi kirjoit
AS			
6	intoja ajatellen tällainen vierailu ei olisi	<i>kovin</i>	hyvä juttu. - Älä kuule, Hujanen
7	saa kokoon. Sitä paitsi en pidä tarinoita	<i>kovin</i>	korkeassa arvossa, kun totuuden
8	yeitekössä. Tavoitteeni ei kuitenkaan ole	<i>kovin</i>	kummoinen . Yritän vain selittää,
9	ta vastapäätä tuvan pöydän luona ja söi.	<i>Kovin</i>	merkillistä ateriaa Hartikainen ei
10	jo liikkeellä. Minusta ne eivät ole edes	<i>kovin</i>	pahoja , vain naurettavia - parhai



Kuvio 7.14. *kovin*: prosodia aineistoittain.

7.2.2.5. *kovin*-määritteen kontekstuaalinen merkitysyksikkö käännössuomessa

Jälleen edeltävästä analyysistä on kerätty taulukkoon 7.17 koonti niistä piirteistä, missä kohdin *kovin*-määritteen käännössuomalainen merkitysyksikkö eroaa alkuperäissuomalaisesta. Kun tiivistelmää verrataan taulukon 7.10 esittämään kuvaan niistä piirteistä, joissa kohdin *hyvin*-määritteen kontekstuaalinen merkitysyksikkö eroaa alkuperäissuomen ja käännössuomen välillä, huomataan *kovin*-määritteen piirreluettelon olevan selvästi pienempi. Erityisesti tämä johtuu KSM- ja AS-aineistojen välisten erojen vähentymisestä, eli *kovin*-määritteen kontekstuaaliset piirteet ovat käännössuomessa ja alkuperäissuomessa samankaltaisempia kuin sen sy-

nonyymilla. Samanlainen tilanne tulee esiin vertailtaessa *hyvin-* ja *kovin-*määritteiden kontekstuaalisten piirteiden eroja AS- ja KSE-aineistojen välillä: nytkin käännös- ja alkuperäissuomen aineistojen välillä on vähemmän eroa *kovin-*määritteen kontekstuaalisissa piirteissä. Tämän havaitsee eritoten semanttista prosodiaa sekä morfologista preferenssiä ja lähes kokonaan myös semanttista preferenssiä erottavien tekijöiden häviämisestä.

Taulukko 7.17. *kovin-*määritteen epätyypilliset kontekstuaaliset kombinaatiot sekä piirteiden yli- ja aliedustuminen käännössuomessa (*suhteessa molempiin muihin aineistoihin).

	KSM-aineisto	KSE-aineisto
Kollokationaalinen valinta	Epätyypilliset kollokaatit* :	Epätyypilliset kollokaatit* :
	<i>helposti</i> , ILOINEN, KAUKANA (adp.), <i>niin</i> , NÄYTTÄÄ, PUHUA, SUURI, <i>tuskin</i> , VAIKEA, VAIKUTTAA, <i>vähän</i>	<i>ehkä</i> , KAUKANA (adv.), <i>kovin</i> , NUORI, PITKÄ, PITKÄLLÄ (adv.), <i>sillä</i> , SURULLINEN
	Epätyypilliset klusterit: (4) <SE EI OLLA <i>kovin</i> >	Epätyypilliset klusterit: (4) <SE EI OLLA <i>kovin</i> >
	Klusteroituminen keskittynyttä	Klusteroituminen keskittynyttä
Kolligationaalinen valinta	–	Aliedustuminen: - substantiivikolligaatti - konjunktiokolligaatti
		Yliedustuminen: - lauseenloppisuuskolligaatti
Morfologinen preferenssi	Yliedustuminen: - liitepartikkeli <i>kaan (kin)</i>	Lyhyt lausepituus
		–
Semanttinen preferenssi	–	Yliedustuminen: - arvo
Semanttinen prosodia	–	–

7.2.3. *oikein*

7.2.3.1. Kollokationaalinen valinta

Taulukko 7.18 luettelee *oikein*-määritteen merkitsevät kollokaatit aineistoittain. Määrite saa kaikissa aineistoissa kaksi samaa tilastollisesti merkitsevää kollokaatiota, näitä ovat

oikein <hyvin, HYVÄ>.

Kollokaatit ovat jokaisessa aineistossa myös erittäin merkitseviä, mistä kertovat suuret merkitsevyysindeksit (taulukko 7.18). Pelkästään käänöskielelle tyypillisiä tilastollisesti merkitseviä kollokaatteja ovat

oikein <MUKAVA, SOPIA>.

Ne eivät ole aivan tuntemattomia AS-aineistossakaan, sillä MUKAVA esiintyy kontekstissa kahdesti ja SOPIA kerran. Vähäisen esiintymistäajuuden vuoksi ne eivät kuitenkaan osoittaudu AS-aineistossa tyypillisiksi kollokaateiksi. Käänössuomen aineistosta poimittuja esimerkkejä on konkordanssissa 7.13.

Taulukko 7.18. Tilastollisesti merkitsevät *oikein*-määritteen kollokaatit osakorpuksittain (vain yhdessä aineistossa merkitsevät kollokaatit kursivoitu).

AS-aineisto			KSM-aineisto			KSE-aineisto		
kollokaatti	MI	t	kollokaatti	MI	t	kollokaatti	MI	t
hyvin	6,29	5,13	hyvin	6,41	6,01	<i>kovasti</i>	6,01	2,95
hyvä	4,46	4,37	mukava	6,03	2,79	hyvin	5,96	5,48
<i>iso</i>	3,56	2,05	<i>hauska</i>	5,82	2,20	<i>kiitos</i>	5,96	3,41
<i>tuntua</i>	3,41	2,03	hyvä	4,98	5,65	mukava	5,03	2,56
<i>jos</i>	2,50	2,47	sopia	4,51	2,53	hyvä	4,63	5,26
			<i>olla</i>	0,61	2,82	sopia	4,48	2,14
						<i>paljon</i>	3,25	2,83

Vaikka tilastollisesti merkitsevien yli viisi kertaa tavattavien kollokaattien määrä on *oikein*-määritteellä vähäinen, on jokaisesta aineistosta löydettävissä sellaisia kollokaatteja, jotka eivät toisissa aineistoissa osoittaudu tyypillisiksi (kursivoidut kollokaatit taulukossa 7.18). Yksikään näistä ei ole sellainen, etteikö sitä tavattaisi *oikein*-määritteen kontekstista myös jossakin toisessa aineistossa. Näiden esiintymät ovat yleensä kuitenkin pieniä, yhdestä kolmeen, joten ne ovat joko yhden kielenkäyttäjän tekstistä peräisin tai muuten harvinaisia. OLLA on näistä poikkeus, sillä se esiintyy taajaan kaikissa aineistoissa (ks. myöhemmin taulukkoa 7.19). Se osoittautuu kuitenkin merkitseväksi ainoastaan KSM-aineistossa, muissa aineistoissa OLLA on kaukana merkitsevyysrajoista. Sen suhteellinen esiintymistäajuus onkin

KSM-aineistossa suurempi kuin muissa aineistoissa (taulukko 7.19). Esiintymistiheyttä selittävät osittain liittomuodot ja kiteytyneet verbiliitot, joissa OLLA-verbi on sijoittunut positioihin O2 ja O3. Tällaisia tavataan KSM-aineistossa (ja myös KSE-aineistossa) kymmenisen kappaletta mutta AS-aineistossa niitä on vain yksi; pienissä esiintymämäärissä tällaisetkin erot ovat merkittäviä. Esimerkkejä tapauksista konkordanssissa 7.14.

Konkordanssi 7.13. Esimerkkejä molemmille käännösaineistoille tyypillisistä kollokaateista.

KSM	
1	eti tunnistanut... Ja emäntä oli miellyttävä. <i>Oikein mukava</i> emäntä! Alkoi heti hössöttä
2	ttä nyt minun pitää taas ryhtyä töihin... Oli <i>oikein mukavaa</i> , hyvä rouva Frühauf... Va
3	a viehättävä naimaikäinen tytär, joka <i>sopii</i> hyvin sinun temperamenttiisi. Myöt
4	aupunkimme väestöräjähdys <i>sopi</i> minulle <i>oikein</i> hyvin. Ihmiset, joilla oli lukematto
KSE	
5	nohtaa, mutta en sinun käsiäsi. Kylpy oli <i>oikein mukava</i> ." Squaw hymyili, mutta h
6	oidan vain haastattelut. Oli hauska tavata. <i>Oikein mukavaa</i> että päätit ottaa tämän as
7	veteen. 'Voi olla, mutta se <i>sopii</i> minulle <i>oikein</i> hyvin.' Jatkoimme matkaa. Grosven
8	aa häntä nimeltä. No, se <i>sopi</i> Kohtuudelle <i>oikein</i> hyvin. Hän melkein raastoi hevosen

Konkordanssi 7.14. Esimerkkejä OLLA-kollokaatista eri aineistoissa.

KSM	
1	haluan sinut. Mikä <i>oli sopinut</i> Jenniferille <i>oikein</i> hyvin. Se oli ollut aika hassua - hän
2	a keittiöntuoleilla. Martha <i>oli olevinaan</i> " <i>oikein</i> kiukkuinen", koska Madeleine oli as
3	Se oli varmaa! Ja joka kerran hän <i>oli ollut</i> <i>oikein</i> pettynyt siitä, ettei isä ollut tavattavi
KSE	
4	sta hauskaa, että tätisi <i>on murhattu</i> ? Että <i>oikein</i> hauskaa, vai? Isäsikin mielestä se oli
5	ä katuvaisena. "Sinä <i>olet hoitanut</i> kaiken <i>oikein</i> hienosti." "Niin olet", sanoi Ręzsi.
6	an luona äiti sanoi, että minun <i>oli oltava</i> <i>oikein</i> kiltti pieni tyttö. Päivisin, kun kaupa
AS	
7	en muun väen. Ja niin <i>olisi</i> kaikki <i>mennyt</i> <i>oikein</i> hyvin. Anneli otti viimeiset kuvat H

Kollokaatioanalyysin perusteella näyttää siltä, että *oikein*-määritteen kotekstien epätyypillisten kollokaatioiden määrä on vähäisempää kuin muilla määritteillä. Aineistot jakavat suhteessa enemmän samoja tilastollisesti merkitseviä kollokaatteja (HYVÄ, *hyvin*), eikä yksikään tilastollisesti merkitsevä kollokaatti ole sellainen, joka ei esiintyisi jossakin toisessa aineistossa. Määritteen yleisimmät kollokaatit on lueteltu taulukossa 7.19; ne ovat aineistoittain verrattain samanlaisia.

Taulukko 7.19. Astemääritteen *oikein* vähintään 5 kertaa esiintyvät kollokaatit yleisyyssjärjestyksessä tarkastelualueella V3–O2. f = frekvenssi, f % = suhteellinen frekvenssi kaikista kollokaateista, F = summafrekvenssi, F % = suhteellinen summafrekvenssi. (Suhteelliset osuudet on laskettu kollokaattien yhteismääristä, jotka ovat AS-aineistossa 605, KSM-aineistossa 785 ja KSE-aineistossa 780.)

	AS-aineisto				KSM-aineisto				KSE-aineisto					
	f	f%	F	F%	f	f%	F	F%	f	f%	F	F%		
olla	29	4,8	29	4,8	olla	67	8,5	67	8,5	olla	57	7,3	57	7,3
hyvin	27	4,5	56	9,3	hyvin	37	4,7	104	13,2	hyvin	31	4,0	88	11,3
hyvä	21	3,5	77	12,7	hyvä	34	4,3	138	17,6	ja	31	4,0	119	15,3
ja	18	3,0	95	15,7	hän	20	2,5	158	20,1	hyvä	30	3,8	149	19,1
se	17	2,8	112	18,7	ja	18	2,3	176	22,4	hän	28	3,6	177	22,7
minä	15	2,5	127	21,0	se	17	2,2	193	24,6	minä	21	2,7	198	25,4
jos	9	1,5	136	22,5	minä	15	1,9	208	26,5	se	21	2,7	219	28,1
ei	8	1,3	144	23,8	sanoa	9	1,1	217	27,6	kiitos	12	1,5	231	29,6
että	7	1,2	151	25,0	kun	8	1,0	225	28,7	ei	11	1,4	242	31,0
mutta	6	1,0	157	26,0	mukava	8	1,0	233	29,7	paljon	10	1,3	252	32,3
sanoa	6	1,0	163	26,9	ei	7	0,9	240	30,6	sanoa	10	1,3	262	33,6
tämä	6	1,0	169	27,9	että	7	0,9	247	31,5	kovasti	9	1,2	271	34,7
hän	5	0,8	174	28,8	sopia	7	0,9	254	32,4	he	7	0,9	278	35,6
iso	5	0,8	179	29,6	voida	7	0,9	261	33,2	mukava	7	0,9	285	36,5
kaikki	5	0,8	184	30,4	jokin	6	0,8	267	34,0	että	6	0,8	291	37,3
nyt	5	0,8	189	31,2	mutta	6	0,8	273	34,8	jokin	6	0,8	297	38,1
sinä	5	0,8	194	32,1	nyt	6	0,8	279	35,5	saada	6	0,8	303	38,8
sitten	5	0,8	199	32,9	te	6	0,8	285	36,3	sinä	6	0,8	309	39,6
tuntua	5	0,8	204	33,7	hauska	5	0,6	290	36,9	tämä	6	0,8	315	40,4
					kaikki	5	0,6	295	37,6	mikä	5	0,6	320	41,0
					me	5	0,6	300	38,2	niin	5	0,6	325	41,7
					paljon	5	0,6	305	38,9	näyttää	5	0,6	330	42,3
										sopia	5	0,6	335	42,9
										tehdä	5	0,6	340	43,6
										tuntua	5	0,6	345	44,2

Leksikaalisia suhteita tarkastellaan seuraavassa vielä klustereiden avulla. Nelisanaisten klusteroitumisasteet ovat AS-aineistossa 3,3 ja käännoaineistoissa 5,1 (KSM) ja 3,8 (KSE) sekä kolmisanaisissa vastaavasti 14,9, 30,6 ja 26,3. Klusteroituminen on arvojen perusteella voimakkaampaa käänno-suomessa riippumatta lähdekielestä; erot kolmisanaisissa klustereissa ovat merkitseviä⁹⁴. Nelisanaissa erot ovat vähäisiä mutta suunnaltaan samanlaisia kuin kolmisanaisissakin. *oikein*-määrittelyllä on siis lähisynonyymejaan selkeämpi taipumus klusteroitua nimenomaan käänno-skielessä, *kovinhan* ei klusteroitunut käänno-suomessa yhtään enempiä kuin alkuperäissuomessa, ja *hyvin* klusteroitui alkuperäissuomea voimakkaammin vain englannista suomennetuissa teksteissä. KSE-aineistossa tavattavaa voimakasta klusteroitumista tukee *oikein*-määrittelyn tapauksessa lisäksi se, että aineiston ainoa viisisanainen klusteri, jossa *oikein* on yhtenä jäsenenä, löytyy tästä aineistosta. Li-

⁹⁴ AS vs. KSM $z = 3,049633$ ($p < 0,01$); AS vs. KSE $z = 2,299718$ ($p < 0,05$).

säksi klusteri jatkuu yli tarkastelualueen aina positioon V4-asti (taulukko 7.20 s. 194).

Vaikka klustereiden leksikko on hyvin samantapainen kaikissa aineistoissa, on erojakin havaittavissa: pronomien variaatio on suurempi käännösaineistoissa, sillä AS-aineistossa ainoa klustereista tavattava pronomini on SE. Myös klustereiden verbit vaihtelevat jonkin verran: AS-aineistossa on OLLA-kollokaatin lisäksi TUNTUA, VOIDA, MUISTAA, KSM-aineistossa uusina TIETÄÄ ja SOPIA sekä KSE-aineistossa TULLA. O1-aseman lekseemit ovat samankaltaisia lukuun ottamatta KSM-aineiston MUKAVA-adjektiivia. Nelisanaisten klustereiden määrä osoittautuu vähäiseksi, aineiston suurempi koko tosin luultavasti lisäisi niiden osuutta ja antaisi mahdollisuuden tarkempaan analyysiin. Näissä pitkissä klustereissa on kuitenkin merkkejä erilaisista rakenteista, vaikka jokaisesta aineistosta löytyykin klusteri <SE OLLA *oikein* HYVÄ>. KSM-aineiston klusterit

<HÄN OLLA *oikein* HYVÄ>

<MINÄ TIETÄÄ *oikein hyvin*>

ja englannista käännettyjen tekstien klusterit

<SOPIA MINÄ *oikein hyvin*>

<TULLA *oikein hyvin* TOIMI>

ovat vieraita alkuperäissuomen aineistolle.

Kolmisanaistenkaan klusterien määrä ei ole suuri verrattuna määritteen synonyymeihin, ja eri klustereiden määrälliset erot ovat pienet. Ainoastaan KSM-aineistossa erottuu yksi kolmisanaisten klusteri selvästi muita yleisemmäksi, rakenne on <OLLA *oikein* HYVÄ>, joka tavataan frekventtinä myös muista aineistoista. KSE-aineiston tavallisin klusteri on <HÄN OLLA *oikein*>, mutta frekvenssiero seuraaviin klustereihin on vain yksi esiintyminen. Kaiken kaikkiaan tämän määritteen klustereissa ilmenevä leksikaalinen vaihtelu näyttää olevan vähäisempää kuin sen synonyymeilla. Klusterianalyysi viittaa edelleen siihen, että *oikein*-määritteen kotekstuaalinen vaihtelu on aineistoittain vähäisempää kuin sen synonyymeilla, tähän tuli jos esiin kollokaatteja tarkasteltaessa. Tämä tarkoittaa sitä, että tämän määritteen alkuperäissuomelle epätyypillisten leksikaalisten suhteiden määrä on sen synonyymeja vähäisempi.

Mahdollista kotekstissa tavattavaa yksinkertaistumista varten on laskettu jälleen frekventeimpien kollokaattien osuudet kaikista tarkastelualueen kollokaateista ja alueella ilmenevä leksikaalinen diversiteetti. Ensinnäkin viiden yleisimmän kollokaatin osuudet ovat käännösaineistossa suuremmat, mutta tulos ei ole merkitsevä 5 prosentin riskitasolla. 10 prosentin riskitaso ylittyy ($z = 1,90 > z_{.10} = 1,64$), joten tulosta voidaan pitää suuntaa antavana. Kymmenen yleisimmän kollokaatin osuus on merkitsevästi suurempi KSE-aineistossa kuin alkuperäissuomessa ja suuntaa-antavasti suurempi kuin yleensä käännössuomessa ($z = 1,66 > z_{.10} = 1,64$). Yleisimpien kollokaattien suuremmasta peitosta käännössuomessa saadaan siis vain suuntaa-antavia tuloksia, joskin englannista käännetty suomennosten suomi osoit-

tautuu erityisesti kymmenen yleisimmän kollokaatin osuuden osalta muista aineistoista poikkeavaksi, eli lähdekielellä tuntuu jälleen olevan vaikutus.

Taulukko 7.20. *oikein*-määritteen klusteroituminen.

	sanoja	V3	V2	V1	noodi	O1	O2	f
AS	4		se	tuntua	<i>oikein</i>	hyvä		2
			se	olla	<i>oikein</i>	hyvä		2
	3			olla	<i>oikein</i>	hyvä		5
			se	olla	<i>oikein</i>			3
				voida	<i>oikein</i>	hyvin		3
				muistaa	<i>oikein</i>	hyvin		3
KSM	4		hän	olla	<i>oikein</i>	hyvä		3
			se	olla	<i>oikein</i>	hyvä		3
			minä	tietää	<i>oikein</i>	hyvin		2
	3			olla	<i>oikein</i>	hyvä		10
				olla	<i>oikein</i>			5
			hän	olla	<i>oikein</i>			4
			se	olla	<i>oikein</i>			4
				olla	<i>oikein</i>	mukava		4
				sopia	<i>oikein</i>	hyvin		4
				voida	<i>oikein</i>	hyvin		4
KSE	(6) 5	(mutta) se	sopia	minä	<i>oikein</i>	hyvin		2
	4		sopia	minä	<i>oikein</i>	hyvin		2
			se	olla	<i>oikein</i>	hyvä		2
				tulla	<i>oikein</i>	hyvin	toimi	2
	3		hän	olla	<i>oikein</i>			6
			se	olla	<i>oikein</i>			5
				olla	<i>oikein</i>	hyvä		5
				<i>oikein</i>	hyvin	hän	4	

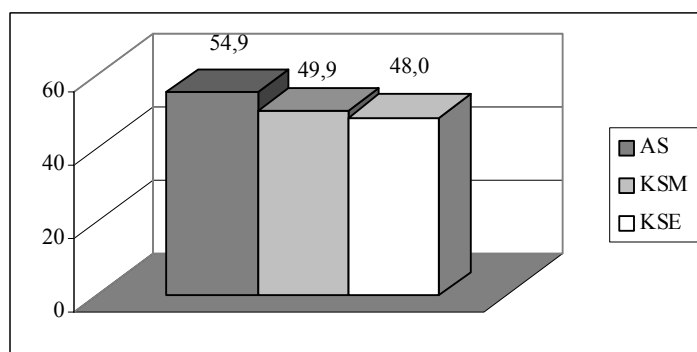
Taulukko 7.21. *oikein*: yleisimpien kollokaattien summafrequenssien suhteellisten osuuksien vertailu.

	AS vs. KSM	AS vs. KSE	KSM vs. KSE
5 yleisimmän osuus	18,5 % vs. 22,4 % z = 1,78, p > 0,05	18,5 % vs. 22,7 % z = 1,90, p > 0,05	22,4 % vs. 22,7 % z = 0,13, p > 0,05
10 yleisimmän osuus ^o	27,9 % vs. 29,7 % z = 0,71, p > 0,05	27,9 % vs. 33,6 % z = 2,25, p < 0,05	29,7 % vs. 33,6 % z = 1,66, p > 0,05

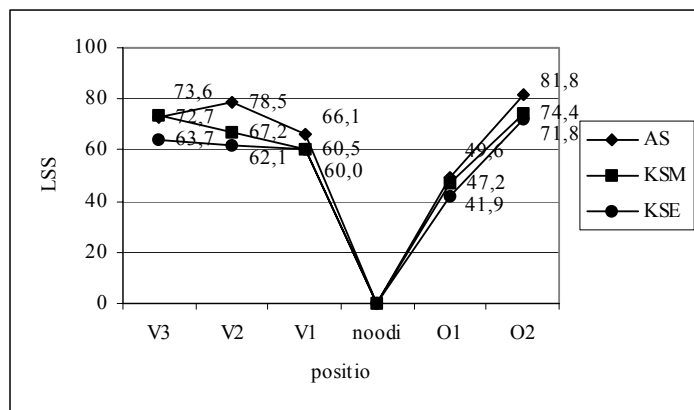
^oKymmenen yleisimmän kollokaatin joukkoon on laskettu AS:ssa 12 ja KSE:ssa 11 kollokaattia, koska myös katkaisukohtaan ylittävillä kollokaateilla on sama frekvenssi.

Koko tarkastelualueen kotekstista laskettu lemman sane-suhde sen sijaan osoittaa leksikaalisen variaation olevan käänösaineistossa pienempi kuin alkuperäisruokien aineistossa (kuva 7.15): tilastollisen testauksen jälkeen AS- ja KSE-aineistojen välinen ero osoittautuu tilastollisesti merkitseväksi ja AS- ja KSM-aineistojen väli-

nen ero suuntaa-antavaksi, käännessaineistojen välillä ei ole merkitsevää eroa⁹⁵. Positiokohtaisen tarkastelun (kuvio 7.16) mukaan kollokationaalinen diversiteetti eri asemilla ei eroa juurikaan aineistoittain: ainoastaan positiossa V2 diversiteetti on molemmissa käännessuomen aineistoissa tilastollisesti merkitsevästi pienempi kuin alkuperäissuomessa⁹⁶. Käännessuomen V2-aseman pienempi kollokationaalinen diversiteetti johtuu KSM:ssä OLLA-kollokaatin (osuus 14,6 %) ja KSE:ssä OLLA- ja HÄN-kollokaattien (osuudet 7,1 % ja 8,3 %) suurista määristä tässä asemassa, AS näiden osuudet ovat selvästi pienemmät (OLLA 1,6 % ja HÄN 2,5 %).



Kuvio 7.15. *oikein*: lemma–sane-suhde tarkastelualueella V3–O2.



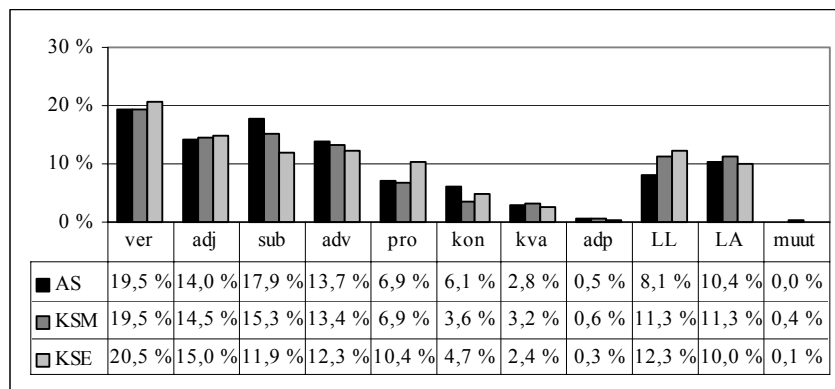
Kuvio 7.16. *oikein*: lemma–sane-suhde positioittain.

⁹⁵ AS vs. KSM $z = 1,739854$ ($p < 0,10$); AS vs. KSE $z = 2,412121$ ($p < 0,05$); KSM vs. KSE $z = 0,678969$ ($p > 0,05$).

⁹⁶ AS vs. KSM $z = 1,992535$ ($p < 0,05$); AS vs. KSE $z = 2,808784$ ($p < 0,01$).

7.2.3.2. Kolligationaalinen valinta

oikein-määritteen synonyymien kolligationaalisen valinnan tarkastelussa on tullut ilmi englannista käännettyjen tekstien ero muiden aineistojen teksteihin. Tämä on näkynyt siinä, että AS- ja KSM-aineistot ovat osoittautuneet hyvin samanlaisiksi niin sanaluokkakolligaattien osuuksien kuin lausepituuksienkin suhteen, KSE-aineisto on taas poikennut näiden kahden muun aineiston ominaisuuksista. Myös *oikein*-määritteen koteksissa kolligaattien osuuksissa on eroa aineistojen välillä. KSE-aineisto poikkeaa tässäkin tapauksessa muista aineistoista, mutta nyt myös KSM-aineiston kolligationaalinen valinta näyttää eroavan alkuperäissuomen preferensseistä, eli *oikein* poikkeaa tässä synonyymeistaan. Edelleen tilanne on kuitenkin se, että kaikki aineistot jakavat samat sanaluokkakolligaatit, eli erot ovat kolligaattien osuuksien eroja ja siten tukevat vain oletusta epätyypillisistä frekvensseistä, eivät epätyypillisistä kombinaatioista. Sanaluokkakolligaattien suhteelliset osuudet ja tilastollisesti merkitsevät erot on esitetty kuviossa 7.17 ja taulukossa 7.22.



Kuvio 7.17. *oikein*: kolligaatit tarkastelualueella V3–O2.

Taulukko 7.22. *oikein*: kolligaatit, joiden osuudet erottavat aineistot toisistaan tilastollisin perustein koko tarkastelualueella.

AS - KSM	AS - KSE	KSM - KSE
	pronominit	
konjunktiot	$z = 2,33, p < 0,05$	
$z = 2,23, p < 0,05$	substantiivit	pronominit
lauseenloppuisuus	$z = 3,11, p < 0,01$	$z = 2,47, p < 0,05$
$z = 2,01, p < 0,05$	lauseenloppuisuus	
	$z = 2,54, p < 0,05$	

Käännössuomen aineistoista KSM:n koteksissa konjunktioiden osuus osoittautuu pienemmäksi ja lauseenloppuisuus suuremmaksi kuin alkuperäissuomessa. AS-aineistossa konjunktioiden osuudet ovat suuremmat V3- ja erityisesti O2-positiossa,

jossa osuukien ero on lähes tilastollisesti merkitsevä⁹⁷. Konjunktiot ovat O2-ase-
massa paria poikkeusta lukuun ottamatta adjektiiveja yhdistäviä rinnastuskonjunkti-
oita sekä V3-asemassa alistus- että rinnastuskonjunktioita. Myös englannista suo-
mennetut tekstit eroavat lauseenloppuisuuden suhteen mutta lisäksi pronominiinien ja
substantiivien osuudet ovat merkitsevästi erilaisia. Pronominien osuus erottaa myös
käännösaineistot toisistaan. KSE-aineiston vähäisen substantiivikolligaattien mää-
rän syy löytyy ainakin osittain position O2 substantiivikollokaateista: alkuperäis-
suomen aineistolle on nimittäin erityisen tyypillistä, että siinä esiintyy usein ra-
kenne <*oikein* ADJ SUB>, jossa siis *oikein* on määritteenä lausekkeessa, jossa sub-
stantiivi on pääsanana (konkordanssissa 7.15 rivit 1–4). Tällaisia rakenteita on pe-
rästi neljännes (24,8 %) kaikista *oikein*-määritteen esiintymistä AS-aineistossa, KSE-
aineistossa osuus on vain 12,8 %. KSE-aineistossa pronominiinien osuus on muiden
aineistojen osuuksia suurempi erityisesti positiossa V3, ja erot ovat myös tilastolli-
sesti merkitseviä⁹⁸. Myös muissa vasemmanpuoleisissa positioissa pronomineja on
hieman enemmän kuin muissa aineistoissa. Englannista käännettyissä teksteissä
vasemmanpuoleinen koteksti on siten pronominiivaltaisempi; hyvin usein pronomini
on HÄN, joka osoittautui jo aiemmin kollokaattianalyyseissa (taulukot 7.28 ja 7.30)
käännöskielelle tyypilliseksi *oikein*-määritteen pronominikollokaatiksi. Näistä tapa-
uksista on esimerkkejä riveillä 5–9.

Konkordanssi 7.15. Esimerkkejä substantiivi- ja pronominikolligaateista KSE-aineistossa.

1	n. Ja vaarin koko elämälle heitin yhen	<i>oikein</i>	<i>ison kiven</i> . Siitä lähtiki kumea jyly.
2	yiseen avioliittoon. Vanhan Kellarissa	<i>oikein</i>	<i>hyvä kaveri</i> pyysi kauniisti puhelinn
3	t erottaa sanojeni alta viiltävä ironia. -	<i>Oikein</i>	<i>hyvää matkaa</i> , minä sanoin. - Pitäkä
4	vuosikertaa. Panevatkohan he joskus	<i>oikein</i>	<i>tunnelmaisena takkailtana</i> kypsän p
5	kan säästöjä, ja jos hän panisi kätensä	<i>oikein</i>	syvälle taskuun, raha saattaisi riittää
6	karille mainiosti. Hän puhui hollantia	<i>oikein</i>	hyvin, melkein yhtä hyvin kuin engla
7	"Odottakaa vain!" sanoi täti. " Hän on	<i>oikein</i>	suloinen tyttö. Tapaamisen arvoinen.
8	lä oli karmaiseva tunne, että hän tiesi	<i>oikein</i>	hyvin. "Ta-Kumsaw", Tack sanoi.
9	aisia ja ymmällään, mikä sopi hänelle	<i>oikein</i>	hyvin. Hän ei kuitenkaan ratkenn

oikein-määritteen kolligaattiklustereita tarkasteltaessa havaittiin (luku 6.2.3), että
yhtä, selvästi muita yleisempää klusteria oli vaikea erottaa, koska selkeästi erottuvia
klusterirakenteita oli koko määritteen esiintymien vähäisyyden vuoksi vain muu-
tama ja eri klusterit erosivat yleisyyksiltään vain vähän. Tyypillisimmäksi, mutta
kuitenkin vain kuusi kertaa esiintyväksi saattoi aineiston perusteella kuitenkin eh-
dottaa klusteria

<?LA SUB VER *oikein* ADJ SUB>,

⁹⁷ AS vs. KSM $z = 1,914701$ ($p < 0,10$).

⁹⁸ AS vs. KSE $z = 1,971542$ ($p < 0,05$); KSM vs. KSE $z = 2,012221$ ($p < 0,05$).

jonka osuus oli 5,0 % kaikista määritteen esiintymistä. KSM-aineistossa tyypillisimmäksi, seitsemän kertaa (4,5 %) tavattavaksi klusteriksi erottuu V2-positioltaan edellisestä erottuva rakenne

<LA PRO VER *oikein* ADJ SUB>.

KSE-aineistosta ei puolestaan löydy yhtä muita yleisempää klusteria vaan molemmat seuraavista rakenteista esiintyvät viisi kertaa (3,2 %):

<PRO VER PRO *oikein* ADV LL>

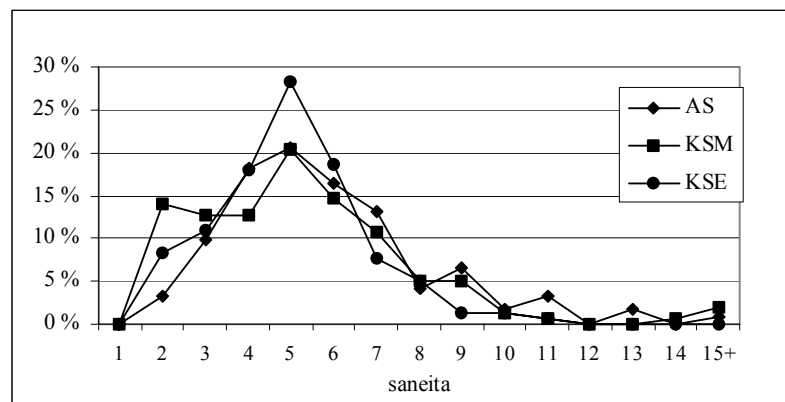
<LA PRO VER *oikein* ADJ LL>.

Käännösaineistot ja eritoten KSE-aineiston erottavaksi ja tyypilliseksi piirteeksi nousee pronominiinien osuus klusterien vasemmanpuoleisissa positioissa ja lauseenloppuisuus O2-positiossa. Tulos on yhtenevä sanaluokkakolligaattianalyysin kanssa, jossa KSE-aineiston kieli poikkesi muista juuri pronomini- ja lauseenloppuisuuskolligaattien suuren määrän vuoksi. Vaikka klustereita tavataankin aineistoittain vähän, alle kymmenen esiintymää, viittaavat aineistojen erilaiset kolligaattiklusterit siihen, että käännöskielen klusterit poikkeavat alkuperäissuomen klustereista ja että englannista suomennettujen tekstien klusterit ovat muihin aineistoihin verrattuna erilaisia. Tulos on viitteellinen, ja se voitaisiin varmistaa kasvattamalla aineistoja niin suuriksi, että jokaisesta aineistosta erottuisivat selvät tyypillisimmät klusterit. Tavallisimpien klustereiden osuuksien erot ovat hyvin pienet, eli klusteroituminen kiinteisiin kolligaattirakenteisiin ei ole käännöskielessä voimakkaampaa kuin alkuperäiskielessäkään. Konkordanssiin 7.16 on lueteltu KSM:n (1–6) ja KSE:n (8–16) frekventeimpien klusterirakenteiden toteutumet.

Konkordanssi 7.16. KSM- ja KSE-aineistojen yleisimpien kolligaattiklustereiden esiintymiä.

| KSM | | | |
|-----|---|---------------|---|
| 1 | Svensson katsoi Annia. Vai niin, <i>tuo oli</i> | <i>oikein</i> | <i>hieno katse</i> , hän sanoi. Julie huoka |
| 2 | una mennä. Mennä yhdessä. <i>Minä tiedän</i> | <i>oikein</i> | <i>huva sauna.</i> Hän mulkoili minua |
| 3 | hoidan tämän omalla tavallani. " <i>Hän on</i> | <i>oikein</i> | <i>mukava nainen</i> , ja -" "Ikävä kyllä |
| 4 | tti?" " <i>Hän oli</i> tyyni, kuten aina. <i>Hän on</i> | <i>oikein</i> | <i>hyvä kapteeni.</i> " Tytön katse harhail |
| 5 | en. Ei hän poikaa rakastanut, <i>hänellä oli</i> | <i>oikein</i> | <i>hyvä suhde</i> mieheensä, jolle hän oli |
| 6 | a", nainen sanoi. "Mitä varten?" " <i>Se on</i> | <i>oikein</i> | <i>hyvä järjestö</i> , hyviä ihmisiä. En |
| KSE | | | |
| 7 | vihaisia ja ymmällään, <i>mikä sopi hänelle</i> | <i>oikein</i> | <i>hyvin.</i> Hän ei kuitenkaan ratkennut |
| 8 | tä", Haarniska sanoi. "Ja <i>se käykin heiltä</i> | <i>oikein</i> | <i>hyvin</i> ", yksi ruotsalaisista sanoi. "P |
| 9 | veteen. "Voi olla, mutta <i>se sopii minulle</i> | <i>oikein</i> | <i>hyvin.</i> Jatkoimme matkaa. Grosve |
| 10 | set saavat rahaa. Mutta <i>se sopii minulle</i> | <i>oikein</i> | <i>hyvin</i> , kunhan en vain ole yksi niis |
| 11 | syyspapeereissaan. " <i>Minä muistan Teidät</i> | <i>oikein</i> | <i>hyvin</i> ", hän sanoi miehelle. "Te käv |
| 12 | valaisevasta kirjeestäsi. <i>Minulle kuuluu</i> | <i>oikein</i> | <i>hyvää</i> ja olen edistynyt huomattav |
| 13 | hän?" "Ei", vastasin minä, " <i>tämä näyttää</i> | <i>oikein</i> | <i>sievältä.</i> Kiitoksia." Kaapiston toise |
| 14 | He söivät yhdessä päivällisen. " <i>Hän on</i> | <i>oikein</i> | <i>hyvännäköinen</i> ", sanoi Ręzsi, "ja |
| 15 | lenhan minä. Niin on äitinikin. <i>Hän on</i> | <i>oikein</i> | <i>lahjakas</i> , ja isän uran alkuaikoina |
| 16 | a ja niin nuori ja niin kauhistunut. <i>Se on</i> | <i>oikein</i> | <i>viehkeää.</i>) Suljin silmäni. Olin hel |

Kuten molempien edellistenkin määritteiden kohdalla, myös *oikein*-määritteen kolligaattianalyyseissä lauseenloppuisuus erottuu merkitseväksi tekijäksi, minkä perusteella voi jälleen epäillä lausepituutta pienemmäksi englannista käännettyissä teksteissä. Tällä kertaa voi myös olettaa, että KSM-aineistonkin lausepituus olisi pienempi kuin alkuperäissuomen aineiston, koska lauseenloppuisuuden osuus on tässäkin aineistossa suurempi. Tilanne onkin hieman erilainen kuin edellä, mutta ei aivan sellainen kuin tässä oletettiin. Englannista käännettyjen tekstien *oikein*-määritteen kontekstuaalinen lausepituus on merkitsevästi pienempi kuin alkuperäissuomessa, mutta ei kuitenkaan pienempi kuin käännoissuomessa yleensä (taulukko 7.23). KSM-aineistossa lausepituus ei ole tilastollisesti merkitsevästi pienempi kuin alkuperäissuomessa; ero ylittää kuitenkin 10 prosentin riskitason ($z = 1,65 > z_{.10} = 1,64$), joten tulosta voidaan pitää suuntaa antavana. Aineiston kasvattaminen voisi mahdollisesti selvittää näiden aineistojen suhteen. Tyypillisin lausepituus on kaikissa aineistoissa 5 sanetta (kuvio 7.18). Pisin lausepituus on siis alkuperäissuomen aineistossa, ja englannista käännettyjen tekstien kontekstilauseet ovat lyhyempiä kuin alkuperäissuomessa. Lisäksi tulokset viittaavat siihen, että käännoiskielen kontekstilauseet olisivat yleensäkin AS-aineiston lauseita lyhyempiä.



Kuvio 7.18. *oikein*: lauseiden sanemäärä.

Taulukko 7.23. *oikein*: lauseiden sanemäärien keskiarvot ja hajonta sekä keskiarvojen erojen tilastollinen merkitsevyys aineistoittain.

| | AS | KSM | KSE |
|------------------------|---------------------------|---------------------------|---------------------------|
| keskiarvo (saneita) | 5,83 | 5,32 | 5,00 |
| vaihteluväli (saneita) | 13 [2–15] | 16 [2–18] | 13 [2–11] |
| | AS–KSM | AS–KSE | KSM–KSE |
| merkitsevyys | $z = 1,65$ ($p > 0,05$) | $z = 3,20$ ($p < 0,01$) | $z = 1,21$ ($p > 0,05$) |

7.2.3.3. Morfologinen preferenssi

Luvussa 6.3.1 kävi ilmi, että *oikein*-määritteellä ei ole synonyymiensa kaltaista paradigmaattista morfologista preferenssiä, joka ilmenisi liitepartikkelin suffiksoitumisena määritteeseen. Tilanne on samankaltainen myös käännoosuomen molemmissa aineistoissa. Näin ollen *oikein*-määritteellä ei ole paradigmaattisessa morfologisessa preferenssissä epätyypillisiä kombinaatioita eikä myöskään epätyypillisiä frekvenssejä.

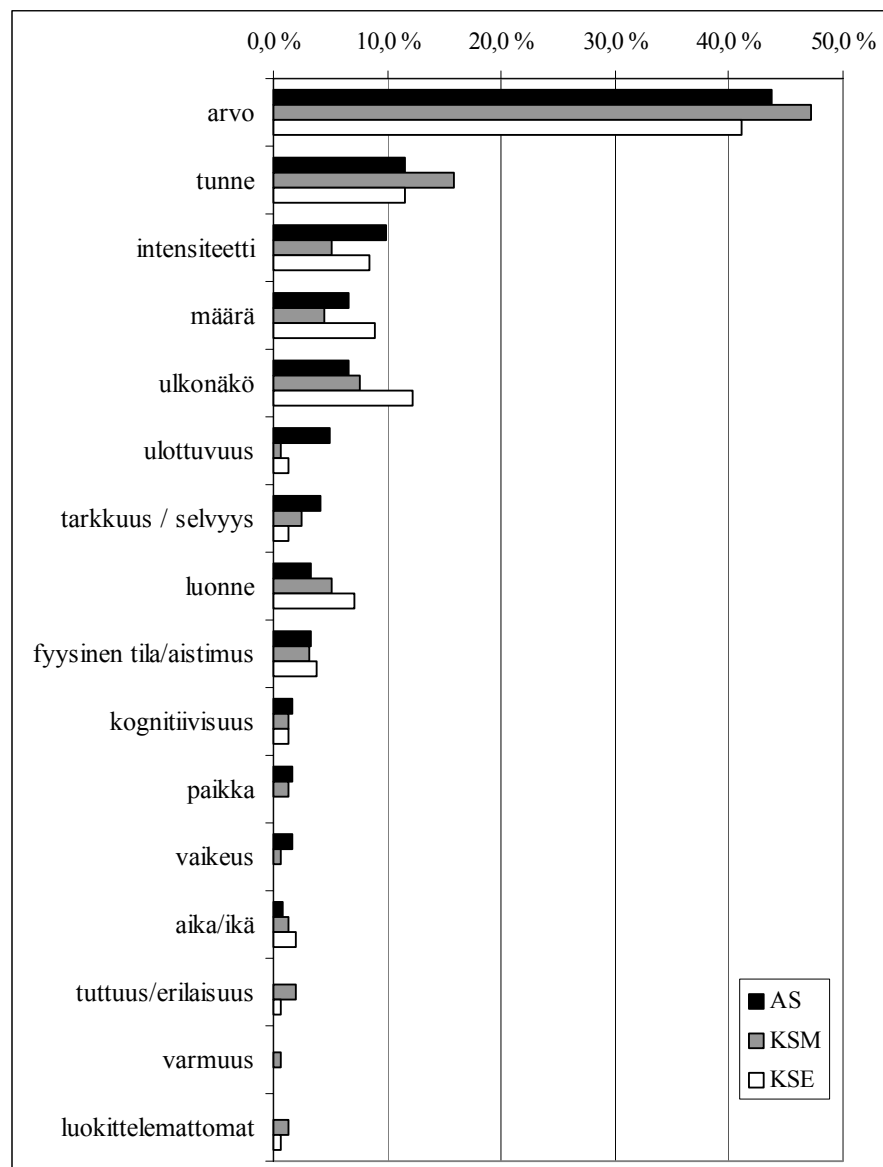
7.2.3.4. Semanttinen preferenssi ja prosodia

Kuvio 7.19 esittää *oikein*-määritteen O1-kollokaattien jakautumisen semanttisiin luokkiin aineistoittain. Suurimmat osuudet ovat arvoa ja tunnetta merkitsevillä luokilla. Edellä *kovin*-määritteen semanttiset preferenssit osoittautuivat lähes toistensa kaltaisiksi kaikkien aineistojen kesken. *oikein* noudattaa tässä *kovin*-synonyymiaan: sen semanttinen preferenssi on hyvin samanlainen kaikissa aineistoissa. Ainoa selkeä, tilastollisesti merkitsevä ero osuuksissa on ulottuvuutta ilmoittavien sanojen osuuksissa. Niiden osuus on käännoiskielessä pienempi kuin alkuperäiskielessä, ja englannista käännettyissä teksteissä suuntaa-antavasti pienempi⁹⁹. Käännettyissä suomenkielisissä teksteissä *oikein* määrittää siten harvemmin ulottuvuutta ilmaisevia sanoja. Muita eroja ei ole näiden kolmen aineiston välille osoitettavissa. Preferenssistä ulottuvuutta merkitseviin sanoihin on esimerkkejä konkordanssissa 7.17.

Konkordanssi 7.17. Esimerkkejä ulottuvuutta ilmaisevista kollokaateista.

| | | |
|---|--|--|
| | AS | |
| 1 | noin rillipäälle, se katto pahasti. Kerätään | <i>oikein</i> iso läjä näitä luovuuista kärsiviä ja |
| 2 | ain vähänkin pyytää. Ja sitten Helena sai | <i>oikein</i> ison lohen. Veijo oli soutamassa ja |
| 3 | itään "vakavaa ja valtavaa", vaan jotakin | <i>oikein</i> pientä ja sokean itsekästä. Jos olisi |
| | KSM | |
| 4 | jon myöhemmin meillä onkin meneillään | <i>oikein</i> iso tappelu, sellainen että Persianla |
| | KSE | |
| 5 | hiukan säästöjä, ja jos hän panisi kätensä | <i>oikein</i> syvälle taskuun, raha saattaisi riitt |
| 6 | ole nähnyt mitään avainta." "Mieli tekee | <i>oikein</i> pitkälle ajelulle. Saamasi pitää." B |

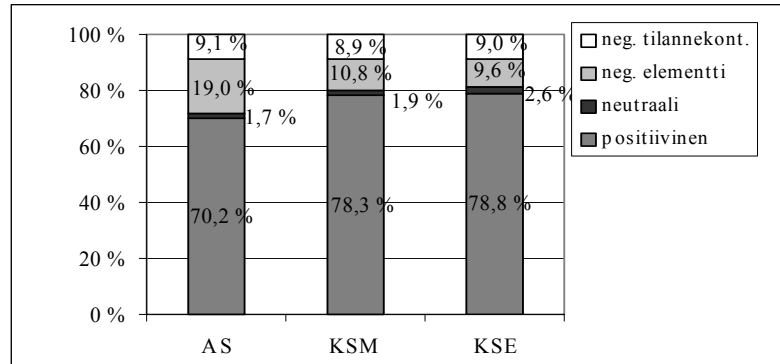
⁹⁹ AS vs. KSM $z = 2,280286$ ($p < 0,05$); AS vs. KSE $z = 1,812271$ ($p < 0,10$).



Kuvio 7.19. *oikein*: semanttinen preferenssi aineistoittain.

Semanttisen prosodian tarkastelu (kuvio 7.20) osoittaa, että käänösaineistoissa prosodia on lähes identtinen, suhdeluvut ovat likimain samoja. Sen sijaan AS-aineiston osuudet poikkeavat näistä kahdesta: positiivisen prosodian osuus on noin 8 prosenttiyksikköä pienempi ja negatiivisen vastaavasti keskimäärin yhdeksän prosenttiyksikköä suurempi. *oikein*-määritteiden vähäisyyden vuoksi näinkään suuret prosenttiyksiköiden erot eivät ole tilastollisesti merkitseviä, joskin erot AS-aineis-

ton ja KSE-aineistojen välillä ovat suuntaa antavia¹⁰⁰. Ainoa tilastollisesti merkitsevä ero tulee esille AS-aineistossa ja KSE-aineistossa tavattavien negatiivisten elementtien osuuksissa. Niiden osuus on englannista käännettyissä teksteissä merkitsevästi pienempi, ja myös KSM-aineistossa suuntaa-antavasti pienempi¹⁰¹. Niiden vähäinen osuus ei kuitenkaan riitä laskemaan koko negatiivisen prosodian osuutta merkitsevästi pienemmäksi.



Kuvio 7.20. *oikein*: prosodia aineistoittain.

Konkordanssi 7.18. Esimerkkejä negatiivisista aineksista *oikein*-määritteen kontekstissa.

| | | |
|-----|---|---|
| AS | | |
| 1 | - Tämä ei nyt vaikuta | <i>oikein</i> hyvältä, Ensio sanoi huolestuneena. - |
| 2 | uskonnon sijaisopettajana erään pienen | <i>oikein</i> ikävä ja hankala oppilas, eräs Jari |
| 3 | kirkonkylän koulussa, niin minulle sattui | Kupiainen, joka puhui lakkaamatta. |
| 4 | "Vaivaako sinua joku, näytät | <i>oikein</i> kurjalta ? - Sain ihottuman kirjasta, |
| 5 | Kuulit oikein | <i>oikein</i> hyvin papatuksen, molinan , kaurapellon kahinastakin pelonhikisen läähätyksen . |
| 6 | - Ai jai, taas pääsi rouva pistämään | <i>oikein</i> pahasti . Mutta totellaan, totellaan tul |
| KSE | | |
| 7 | Ja jos oikein | <i>oikein</i> huonosti kävisi, jos isäntänä ei olisi-kaan Neuvostoliitto, vaan joku paikall- |
| 8 | nahkaiset kesäkengät, joita olin niin usein | <i>oikein</i> hyvin. Päähäni panin Nancy'n hyvän olkihatun |
| 9 | ihailut, vaikka ne eivät sopineetkaan | <i>oikein</i> pahaa valkoisille - enkä minä sano että he semmoista tahtoisivat |
| 10 | "Mutta vaikka he tahtoisivat tehdä jotain | <i>oikein</i> kivaakin saa tehdä, mutta kuka siitä saisi tietää? |
| 11 | "Ei tietenkään, tyhmä. Ei mitään | <i>oikein</i> kivaakin saa tehdä, mutta kuka siitä saisi tietää? |
| 12 | He pysähtyivät leikkaamaan palan pojan ihosta tai puhkaisemaan silmän, | <i>oikein</i> kovasti ja pani pojan huutamaan . |
| 13 | jotain mikä sattui | |

¹⁰⁰ Positiivinen: AS vs. KSM $z = 1,541898$ ($p > 0,05$); AS vs. KSE $z = 1,641146$ ($p < 0,10$).

Negatiivinen: AS vs. KSM $z = 1,631587$ ($p > 0,05$); AS vs. KSE $z = 1,872717$ ($p < 0,10$).

¹⁰¹ KSE vs. AS $z = 2,253735$ ($p < 0,05$), KSM vs. AS $z = 1,926693$ ($p < 0,10$).

7.2.3.5. *oikein*-määritteen kontekstuaalinen merkitysyksikkö käännessuomessa

Myös *oikein*-määritteen konteksti on sen synonyymien *hyvin* kontekstia jähmeämpi muuttumaan käännettäessä. Erottavia piirteitä löytyy kollokationaalisesta valinnasta kuten sen molemmilla synonyymeilla, ja myös kolligationaalisesta valinnasta. Muiden piirteiden muuttuminen onkin sitten lähes olematonta, ja niiden osalta *oikein*-määritteen konteksti on tässä tarkastelussa mukana olevien piirteiden osalta lähes samanlainen, olipa aineisto sitten mikä tahansa. Käännesskielestä tavattavat epätyypilliset kombinaatiot ja piirteiden yli- ja aliedustumiset on esitetty tiivistetysti taulukossa 7.24.

Taulukko 7.24. *oikein*-määritteen epätyypilliset kontekstuaaliset kombinaatiot sekä piirteiden yli- ja aliedustuminen käännessuomessa (*suhteessa molempiin muihin aineistoihin).

| | KSM-aineisto | KSE-aineisto |
|-----------------------------------|--|--|
| | Epätyypilliset kollokaatit*: | Epätyypilliset kollokaatit*: |
| | HAUSKA, OLLA | KIITOS, kovasti, paljon |
| | Epätyypilliset klusterit:
(4) <HÄN OLLA <i>oikein</i> HYVÄ>
<MINÄ TIETÄÄ <i>oikein hyvin</i> > | Epätyypilliset klusterit:
(4) <SOPIA MINÄ <i>oikein hyvin</i> >
<TULLA <i>oikein hyvin</i> TOIMI>
(3) <HÄN OLLA <i>oikein</i> > |
| Kollokationaalinen valinta | ----- | Yliedustuminen:
- 10 yleisintä kollokaattia |
| | – | Kollokationaalinen diversiteetti pienempi |
| | Aliedustuminen:
- konjunktio-kolligaatti | Aliedustuminen:
- substantiivikolligaatti |
| Kolligationaalinen valinta | Yliedustuminen
- lauseenloppuiskolligaatti | Yliedustuminen:
- pronomini-kolligaatti
- lauseenloppuiskolligaatti |
| | | Lyhyt lausepituus |
| Morfologinen preferenssi | – | – |
| Semanttinen preferenssi | Aliedustuminen:
- ulottuvuus | – |
| Semanttinen prosodia | – | (Aliedustuminen:
- negatiivinen elementti) |

7.3. Synonymian muuttuminen käännösuomessa

Aiemmassa tutkimuskirjallisuudessa (ks. lukua 4.3) on esitetty, että synonymymisten ilmausten epätyypillisten frekvenssien ja epätyypillisen kontekstuaalisen valinnan lisäksi itse synonymisyys olisi käännöskielessä erilaista kuin alkuperäiskielessä. On ehdotettu (Kohn 1996), että synonymia yksinkertaistuisi käännöskielessä, mikä osaltaan aiheuttaisi käännöskielen leksikaalista yksinkertaistumista. Ajatuksena on ollut, että kääntämisessä synonymiapesyeen jäsenistä käytettäisiin vain osaa, niiden frekventeimpiä lekseemejä, mutta tätä ei ole vielä todistettu laajoilla aineistoilla vaan jopa vastakkaisia havaintoja on tehty (Jantunen 2001a).

Jotta synonymymiset ilmaukset olisivat korvattavissa toisillaan, olisi luultavaa, että niiden kontekstuaaliset ominaisuudet olisivat muuttuneet samankaltaisemmiksi kuin ne ovat alkuperäissuomessa. Tämä oletus perustuu siihen, että (synonymiisilla) lekseemeillä on tietynlaiset kielen eri osa-alueilla havaittavat kontekstia koskevat valintakriteerit, jotka erottelevat sanat toisistaan, kuten kävi ilmi astemääritteiden osalta luvussa 6. Tämän vuoksi lekseemit ovat korvattavissa vain joissakin tapauksissa toisillaan, mutta eivät aina ja kaikessa kielenkäytössä. Siten synonymian kaventuminen käännöskielessä tuntuu ainakin jossain määrin problematiselta ajatukselta. Edellä ilmeni, että synonymymisten ilmausten valintapiirteet voivat olla käännöskielessä jossain määrin erilaisia kuin alkuperäiskielessä. Analyysi kuitenkin osoitti, että synonymymisten lekseemien kontekstuaalinen erilaisuus säilyy käännöskielessäkin eli ne ovat tässäkin kielimuodossa selvästi toistaan poikkeavia kontekstuaalisia leksikaalisia yksiköitä.

Tutkimukseni viimeinen käännösuomea käsittelevä analyysi koskee lekseemien synonymisuuden muuttumista käännöskielessä. Tavoitteena on tutkia, lähentyvätkö synonymit toisiaan käännöskielessä eli tasoittuvatko niiden väliset kontekstuaaliset erot, ja olisivatko ne siten periaatteellisella tasolla helpommin toisillaan korvautuvia. Analyysissa käytetään hyväksi tämän luvun aiemmissa analyyseissa selvitettyjä kollokationaalisia, kolligationaalisia, morfologisia ja semanttisia preferenssejä, joiden välisiä lukumäärä- ja suhde-eroja on mahdollisuus verrata luvussa 6 esitettyihin tuloksiin. Mittauksessa käytetään hyväksi erottavien tekijöiden lukumäärää, prosenttiosuuksien yksikköeroja sekä tilastollista merkitsevyyttä kuvaavia arvoja; analyysi etenee edelliseen tapaan kollokaatioista semanttiseen prosodiaan.

7.3.1. Kollokaatiot

Synonymymisten lekseemien kollokationaalista eroa alkuperäis- ja käännösuomen kesken on tarkasteltu seuraavassa yhteisten tilastollisten kollokaattien määrän, kollokationaalisen diversiteetin ja klusteroitumisasteen avulla. Tilastollisesti merkitsevät kollokaatit, kollokationaalinen variaatio tarkastelualueella ja astemääritteiden taipumus klusteroitumiseen nousivat määritteitä erotteleviksi piirteiksi, ja kunkin piirteen variaatiota on myös vaivatonta vertailla aineistojen kesken.

AS-aineistossa yhteisten tilastollisten kollokaattien määrä osoittautui vähäiseksi: synonyymipareista *hyvin* ja *kovin* sekä *hyvin* ja *oikein* molemmissa määritteet jakoivat yhden tilastollisesti merkitsevän kollokaatin, mutta määritteet *kovin* ja *oikein* eivät saaneet ainuttakaan yhteistä merkitsevää kollokaattia (taulukko 6.4 s. 89). Yhteisten merkitsevien kollokaattien perusteella astemääritteet eivät osoittautuneet alkuperäissuomen aineistossa siis kovinkaan synonyymisiksi. Seuraavaan taulukkoon (7.25) koottu aineistoittain astemääritteiden yhteiset merkitsevät kollokaatit.

Taulukko 7.25. Yhteiset tilastollisesti merkitsevät kollokaatit astemäärittepareittain ja aineistoittain.

| | <i>hyvin vs. kovin</i> | <i>hyvin vs. oikein</i> | <i>kovin vs. oikein</i> |
|---------------------|------------------------|-------------------------|-------------------------|
| AS-aineisto | olla | hyvin | – |
| | olla | olla | hyvä |
| | paljon
pitkään | | olla |
| KSM-aineisto | suuri | | |
| | tärkeä | | |
| | vaikea | | |
| | vaikuttaa
vähän | | |
| KSE-aineisto | nuori | – | hyvin |
| | olla | | hyvä |
| | paha | | paljon |
| | pitkä | | |
| | surullinen | | |

Tulosten valossa näyttää siltä, että astemääritteparien synonyymisyys ei muodosta yhteistä linjaa: *hyvin*- ja *kovin*-määritteiden sekä *kovin*- ja *oikein*-määritteiden yhteisten merkitsevien kollokaattien määrä on suurempi käännsaineistoissa kuin alkuperäissuomen aineistoissa, *hyvin*- ja *oikein*-määritteillä muutos on vähäinen. Kahden ensin mainitun astemääritteparin jäsenten välinen synonymia näyttää voimistuvan käännsuomessa, koska ne jakavat useampia yhteisiä merkitseviä kollokaatteja. *hyvin*- ja *oikein*-määritteiden keskinäinen kollokationaalinen merkitysero sen sijaan näyttää pysyvän lähes samana eri aineistoissa.

Koska tilastollisesti merkitsevät kollokaatit kattavat vain suppean osan kaikista tarkastelualueilla olevista kollokaateista, olen ottanut huomioon vielä kaikkien vähintään viisi kertaa esiintyvien eli kaikkien kollokaatioanalyysiin päässeiden kollokaattien väliset suhteet. Seuraavassa on lueteltu ne kollokaatit, jotka yhdistävät *kaikki* astemääritteet, ja laskettu niiden osuus kaikista kollokaateista kolmen astemääritteen tarkastelualueilla.

AS-aineisto

hyvin, kovin, oikein <EI, että, HÄN, ja, MINÄ, mutta, OLLA, SE, TÄMÄ; 25,0 %>

KSM-aineisto

hyvin, kovin, oikein <EI, *että* HÄN, *ja*, JOKIN, *kun*, ME, MINÄ, *mutta*, nyt OLLA, *paljon*, SANOA, SE; 27,4 %>

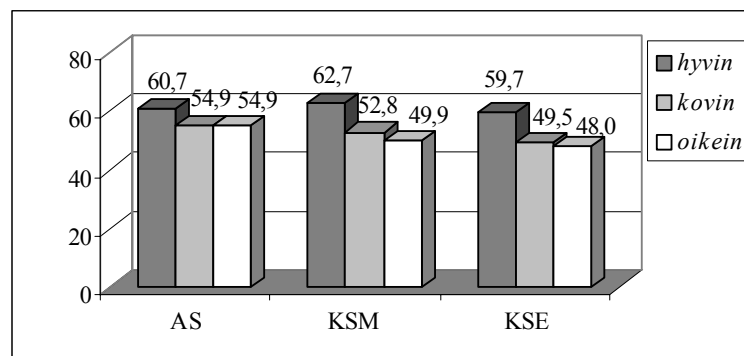
KSE-aineisto

hyvin, kovin, oikein <EI, *että*, HE, HÄN, *ja*, MINÄ, *niin*, NÄYTTÄÄ, OLLA, *paljon*, SANOA, SE, SINÄ, TUNTUA, TÄMÄ; 30,8 %>

Ensinnäkin on huomattavissa, että määritteitä yhdistävien kollokaattien määrä kasvaa käännskielessä: KSM-aineistossa niitä on 13, KSE-aineistossa 15, mutta AS-aineistossa vain 9. Näin ollen määritteillä on enemmän sekä yhteisiä tilastollisesti merkitseviä että suhteellisen frekventtejä, vähintään viisi kertaa kontekstissa esiintyviä kollokaatteja. Alkuperäissuomen aineiston 25 prosentin osuus kertoo lisäksi, että 75 prosenttia tarkastelualueilla esiintyvistä kollokaateista on joitakin muita kuin edellä lueteltuja kaikkia määritteitä yhdistäviä leksemejä. Käännsaineistoissa yhteisten kollokaattien osuus on suurempi, 27,4 % ja 30,8 %, ja siten määritteitä toisistaan erottavien, synonyymisyyttä heikentävien kollokaattien määrä jää pienemmäksi; ero on merkitsevä molempien käännsaineistojen suhteessa AS-aineistoon¹⁰². Tulos on yhdenmukainen edellä tarkastellun yhteisten tilastollisten kollokaattien osoittaman linjan kanssa: käännsuomessa yhteisten kollokaattien määrät ja osuudet ovat suurempia kuin alkuperäissuomessa. Määritteiden keskinäisen synonyymisyyden kannalta tämä tarkoittaa synonyymisyyden voimistumista käännsuomessa, sillä yhteisten kollokaattien suurempi osuus käännsuomessa jättää vähemmän tilaa sellaisille kollokaateille, jotka erottavat synonyymiset määritteet toisistaan.

Tarkastelualueen kollokaattien kokonaisvariaatiota on kuitenkin syytä tarkastella vielä kollokationaalisen diversiteetin avulla. Kuten astemääritteiden synonyymian tarkastelussa kävi ilmi (luku 6.1.2), kollokationaalinen diversiteetti erotteli määritteet toisistaan alkuperäissuomen aineistossa siten, että *hyvin*-määritteen diversiteetti oli *kovin*- ja *oikein*-määritteitä suurempi ja jälkimmäisten diversiteetti yhtä suuri. Vastaava diversiteetin laskenta käännsaineistoista osoittaa, että kollokaattien variaatio on edelleen synonyymeja erottava tekijä. Tilanne kuitenkin muuttuu käännsaineistoissa siten, että *kovin*- ja *oikein*-määritteidenkin välillä on ero kollokationaalisisessa variaatiossa, kuten kuvio 7.21 osoittaa. Astemääriteparien diversiteetin erojen muuttumista voidaan tarkastella merkitsevyyslukujen avulla siten, että jos merkitsevyysarvo eli diversiteetin eron merkitsevyyttä osoittava arvo on muuttunut suuremmaksi, merkitsee se synonyymisyyden vähenemistä määritteiden kesken. Vertailu on esitetty taulukossa 7.26.

¹⁰² AS vs. KSM $z = 2,486454$ ($p < 0,05$); AS vs. KSE $z = 5,652058$ ($p < 0,001$).



Kuvio 7.21. Kollokationaalinen diversiteetti aineistoittain.

Taulukko 7.26. Leksikaalisen diversiteetin merkitsevyyserot astemääriteparien välillä aineistoittain.

| | <i>hyvin vs. kovin</i> | <i>hyvin vs. oikein</i> | <i>kovin vs. oikein</i> |
|---------------------|----------------------------|----------------------------|---------------------------|
| AS-aineisto | $z = 2,00$ ($p < 0,05$) | $z = 2,03$ ($p < 0,05$) | $z = 0,01$ ($p > 0,05$) |
| KSM-aineisto | $z = 3,55$ ($p < 0,001$) | $z = 4,56$ ($p < 0,001$) | $z = 1,02$ ($p > 0,05$) |
| KSE-aineisto | $z = 3,60$ ($p < 0,001$) | $z = 4,14$ ($p < 0,001$) | $z = 0,54$ ($p > 0,05$) |

Astemääriteparien välisten merkitsevyyсарvojen kasvaminen käänösaineistoissa kertoo synonyymisuuden vähentymisestä. Leksikaalisen diversiteetin erot kasvavat aineistojen välillä käänöskielessä kauttaaltaan, myös *kovin*- ja *oikein*-määritteiden välillä, vaikka ero ei muutukaan tilastollisesti merkitseväksi. Arvojen eron kasvaminen on osoitus synonyymisten ilmausten erilaistumisesta käänöskielessä, eli niiden väliset erot ovat käännetyissä kielessä suuremmat kuin alkuperäiskielessä. Englannista käännetyt tekstit antavat samansuuntaisen tuloksen, joten suuntaus ei ole lähdekieliriippuvainen. Tulos on vastakkainen edellä kerrotun kanssa. Kaikille määritteille yhteisiin kollokaatteihin perustuvan analyysin mukaanhan määritteiden synonymia näyttää voimistuvan käänöskielessä, mutta leksikaalinen diversiteetti osoittaa synonyymian laimentumista. Tulokset eivät ole kuitenkaan välttämättä ristiriidassa keskenään. Vaikka yhteisten merkitsevien ja frekventtien kollokaattien määrän lisääntyminen käänösaineistossa osoittaaakin lähentymistä, ei näiden muuttujien avulla pysty päättelemään kotekstin lekseemien kokonaisvariaatiota, johon vaikuttavat lisäavästi harvakseltaan esiintyvät kollokaatit, kuten hapax legomenat.

Myös astemääritteiden taipumus klusteroitumiseen kahta kollokaattia pitempiin leksikaalisiin ketjuihin erotteli määritteen alkuperäissuomessa selvästi. Onko ero

sitten yhtä selvä muissa aineistoissa? Sitä kuvataan taulukossa 7.27, johon on lueteltu frekventimmät neli- ja kolmisanaiset klusterit ja klusteroitumisasteet.

Taulukko 7.27. Klusterit ja klusteroitumisasteet aineistoittain.

| | | <i>hyvin</i> | <i>kovin</i> | <i>oikein</i> |
|------------|---------------|--------------------------------|----------------------------|---------------------------|
| AS | nelisanainen | <että MINÄ OLLA <i>hyvin</i> > | <HÄN EI OLLA <i>ko-</i> | <SE TUNTUA <i>oikein</i> |
| | - klust.aste | 0,6 | <i>vin</i> > | HYVÄ> |
| | kolmisanainen | <SE OLLA <i>hyvin</i> > | 6,3 | 3,3 |
| | - klust.aste | 23,0 | <EI OLLA <i>kovin</i> > | <OLLA <i>oikein</i> HYVÄ> |
| | | | 38,6 | 14,9 |
| KSM | nelisanainen | <MINÄ OLLA OLLA <i>hyvin</i> > | <EI OLLA OLLA | <HÄN OLLA <i>oikein</i> |
| | - klust.aste | 2,1 | <i>kovin</i> > | HYVÄ> |
| | kolmisanainen | <HÄN OLLA <i>hyvin</i> > | 9,7 | 5,1 |
| | - klust.aste | 22,7 | <EI OLLA <i>kovin</i> > | <OLLA <i>oikein</i> HYVÄ> |
| | | | 35,5 | 30,6 |
| KSE | nelisanainen | <MINÄ SE OLLA <i>hyvin</i> > | <SE EI OLLA <i>kovin</i> > | <SOPIA MINÄ <i>oikein</i> |
| | - klust.aste | 3,4 | 9,8 | <i>hyvin</i> > |
| | kolmisanainen | <SE OLLA <i>hyvin</i> > | <EI OLLA <i>kovin</i> > | 3,9 |
| | - klust.aste | 30,3 | 40,8 | <HÄN OLLA <i>oikein</i> > |
| | | | | 26,3 |

Astemääritteiden klusteroituminen säilyy tavallisimpien neli- ja kolmisanaisien klustereiden tarkastelussa erilaisena, sillä klusterit ovat määritteiden kesken toisistaan poikkeavia kaikissa aineistoissa. Käännössuomessa ei siten ole hävinnyt määritteiden välinen kvalitatiivinen ero. Myös astemääritteiden erilainen taipumus klusteroitumiseen näyttää säilyvän käännösaineistoissa, mistä ovat osoituksena määritteiden erisuuret klusteroitumisasteet kaikissa aineistoissa. *hyvin*- ja *kovin*-määritteiden ja *kovin*- ja *oikein*-määritteiden keskinäinen erilainen taipumus nelisanaisiin klustereihin näyttää jopa hieman kasvavan käännössuomessa, sillä klusteroitumisasteiden erotukset ovat käännösaineistossa hieman suuremmat, esimerkiksi AS-aineistossa erotus *kovin*- ja *oikein*-määritteiden välillä on 3,0 yksikköä (6,3 - 3,3), mutta käännösaineistoissa 4,6 (9,7 - 5,1) ja 5,9 (9,8 - 3,9) yksikköä. Ero siis kasvaa määritteiden välillä käännösaineistoissa.

Sen sijaan erot taipumuksessa klusteroitua kolmisanaisiin jaksoihin osoittavat päinvastaista suuntaa, mikä näkyy klusteroitumisasteiden selvänä pienenemisenä käännösaineistoissa. Erimerkiksi *hyvin*- ja *kovin*-määritteiden välinen 15,6 yksikön ero (38,6 - 23,0) kaventuu KSE-aineiston 10,5 yksikön eroon (40,8 - 30,3), ja *hyvin*- ja *oikein*-määritteiden 8,1 yksikön (23,0 - 14,9) ero 4,0 yksikön (30,3 - 26,3) eroon. Rajuinta muutos on KSM-aineistossa *kovin*- ja *oikein*-määritteiden välillä, jossa klusteroitumisherkkyuden ero on kaventunut 4,9 yksikköön (35,5 - 30,6), AS-aineistossa vastaava erotus on peräti 23,7 yksikköä (38,6 - 14,9). Koska kolmisanaiset klusterit ovat huomattavasti frekventimpiä kuin harvinaiset nelisanaiset, voidaan niiden antamaa tulosta pitää tässä yhteydessä tarkempänä. Sen mukaan astemäärit-

teiden välinen ero taipumuksessa klusteroitumiseen on kaventunut käänösuo-
suomessa, eli synonymia näyttää näiltä osin voimistuvan, vaikka klustereiden ra-
kenteelliset erot säilyvätkin.

7.3.2. Kolligaatiot

Myös kolligationaalinen valinta erottelee määritteet toisistaan alkuperäissuomessa:
sanaluokkakolligaattien osuuksissa, kotekstilausepituuksissa ja kolligaattikluste-
reissa oli eroja synonyymisten astemääritteiden kesken. Seuraavaksi näitä ominai-
suuksia käsitellään käänöskorpuksissa, ja kuten edelläkin, synonyymisyyttä verra-
taan AS-aineistossa ilmenneseen. Taulukossa 7.28 on esitetty ensiksi astemäärit-
teitä erottavat kolligaatit ja tilastollista merkitsevyyttä osoittava arvo. Erittäin mer-
kitsevästi ($z > z_{.001} = 3,29$) eroavat kolligaatit on tummennettu.

Taulukko 7.28. Määritteitä erottavat kolligaattipreferenssit aineistoittain.

| <i>hyvin vs. kovin</i> | | <i>hyvin vs. oikein</i> | | <i>kovin vs. oikein</i> | |
|------------------------|------------------------|-------------------------|------------------------|-------------------------|------------------------|
| AS-aineisto | | | | | |
| verbit | $z=6,78$ ($p<0,001$) | LA | $z=2,54$ ($p<0,05$) | verbit | $z=4,70$ ($p<0,001$) |
| LA | $z=2,62$ ($p<0,01$) | adjektiivit | $z=2,01$ ($p<0,05$) | LA | $z=4,37$ ($p<0,001$) |
| adjektiivit | $z=2,36$ ($p<0,05$) | | | konjunktiot | $z=2,10$ ($p<0,05$) |
| konjunktiot | $z=2,22$ ($p<0,05$) | | | | |
| KSM-aineisto | | | | | |
| verbit | $z=8,02$ ($p<0,001$) | LA | $z=5,12$ ($p<0,001$) | LA | $z=5,64$ ($p<0,001$) |
| substantiivit | $z=4,78$ ($p<0,001$) | LL | $z=3,01$ ($p<0,01$) | verbit | $z=4,58$ ($p<0,001$) |
| adjektiivit | $z=3,43$ ($p<0,001$) | adjektiivit | $z=2,18$ ($p<0,05$) | | |
| konjunktiot | $z=2,27$ ($p<0,05$) | substantiivit | $z=2,06$ ($p<0,05$) | | |
| adpositiot | $z=2,18$ ($p<0,05$) | adpositiot | $z=1,98$ ($p<0,05$) | | |
| | | konjunktiot | $z=1,96$ ($p<0,05$) | | |
| KSE-aineisto | | | | | |
| verbit | $z=6,19$ ($p<0,001$) | LL | $z=3,90$ ($p<0,001$) | LA | $z=6,91$ ($p<0,001$) |
| substantiivit | $z=5,02$ ($p<0,001$) | adverbit | $z=3,69$ ($p<0,001$) | pronominit | $z=4,89$ ($p<0,001$) |
| konjunktiot | $z=3,30$ ($p<0,001$) | LA | $z=3,67$ ($p<0,001$) | konjunktiot | $z=4,19$ ($p<0,001$) |
| LA | $z=4,52$ ($p<0,001$) | verbit | $z=3,52$ ($p<0,001$) | substantiivit | $z=2,87$ ($p<0,01$) |
| adjektiivit | $z=3,62$ ($p<0,001$) | pronominit | $z=3,47$ ($p<0,001$) | LL | $z=2,73$ ($p<0,01$) |
| kvanttorit | $z=2,74$ ($p<0,01$) | | | adjektiivit | $z=2,65$ ($p<0,01$) |
| pronominit | $z=2,29$ ($p<0,05$) | | | adverbit | $z=2,55$ ($p<0,05$) |

Taulukosta on nähtävissä kaksi suuntausta. Ensiksikin niiden kolligaattien määrät,
joiden osuudet erottavat määritteet toisistaan, ovat AS-aineiston määriä suurempia
molemmassa käänösaineistoissa. Tämä koskee kaikkia muita vertailuja, paitsi *ko-
vin-* ja *oikein-*määritteitä KSM-aineistossa. Toinen selvä ero aineistojen välillä on
erittäin merkitsevää tilastollista eroa osoittavien kolligaattien määrien ero käänösai-
aineistojen ja alkuperäissuomen aineistojen välillä; näidenkin määrä on käänösai-

neistossa suurempi kuin AS-aineistossa. Molemmat seikat osoittavat, että määritteiden kontekstuaaliset merkityserot kasvavat käännessuomessa ja siten niiden synonyymisyys heikkenee.

Lausepituuden merkitystä tarkastellaan taulukossa 7.29 esitettyjen lausepituuksien keskiarvojen erotuksen avulla. Näistä ei sanaluokkakolligaattien tapaan muodostu yhtenäistä linjaa, vaan synonyymisuuden aste vaihtelee astemäärittepareittain. *hyvin-* ja *kovin-*määritteiden synonyymisyys näyttää ensinnäkin vahvistuvan, koska keskiarvojen erotus on pienempi (0,21 ja 0,19 saneita) käännessuomessa kuin alkuperäissuomessa (0,60 saneita). *hyvin-* ja *oikein-*määritteiden välinen ero näyttää kasvavan käännessuomessa, mutta englannista käännettyissä teksteissä ero on taantunut AS-aineiston tasolle. Viimeisen parin osalta määritteiden ero näyttää jälleen käännessuomessa kasvavan, mutta KSE-aineistossa vuorostaan vähenevän. Lausepituus ei siten osoita, että määritteiden synonymia selvästi vahvistuisi tai heikkenisi käännesskielessä. Pikemminkin analyysi tuo esiin astemääritteiden ja lähdekielen vaikutuksen tuloksiin.

Taulukko 7.29. Lausepituuksien erot määrittepareittain ja aineistoittain.

| | | <i>hyvin</i> | <i>kovin</i> | <i>oikein</i> |
|------------|-------------------------------|--------------------------------|---------------------------------|---------------------------------|
| | keskiarvo (saneita) | 6,68 | 7,28 | 5,83 |
| AS | keskiarvojen erotus (saneita) | <i>hyvin vs. kovin</i>
0,60 | <i>hyvin vs. oikein</i>
0,85 | <i>kovin vs. oikein</i>
1,45 |
| | merkitsevyys | $z=2,40$ ($p<0,05$) | $z=3,19$ ($p<0,01$) | $z=4,87$ ($p<0,001$) |
| | | | | |
| | | <i>hyvin</i> | <i>kovin</i> | <i>oikein</i> |
| | keskiarvo (saneita) | 6,90 | 7,11 | 5,32 |
| KSM | keskiarvojen erotus (saneita) | <i>hyvin vs. kovin</i>
0,21 | <i>hyvin vs. oikein</i>
1,58 | <i>kovin vs. oikein</i>
1,79 |
| | merkitsevyys | $z=1,07$ ($p>0,05$) | $z=6,23$ ($p<0,001$) | $z=6,61$ ($p<0,001$) |
| | | | | |
| | | <i>hyvin</i> | <i>kovin</i> | <i>oikein</i> |
| | keskiarvo (saneita) | 5,82 | 6,01 | 5,00 |
| KSE | keskiarvojen erotus (saneita) | <i>hyvin vs. kovin</i>
0,19 | <i>hyvin vs. oikein</i>
0,82 | <i>kovin vs. oikein</i>
1,01 |
| | merkitsevyys | $z=1,06$ ($p>0,05$) | $z=4,73$ ($p<0,001$) | $z=5,25$ ($p<0,001$) |
| | | | | |

Myös kolligaattiklusterit olivat varsin erilaisia määritteiden kesken. Viimeisenä kolligaatteihin liittyvänä ominaisuutena tarkastellaankin kolligaattien klusteroitumista käännessaineistoissa. Klusterit ovat aineistoittain seuraavat, tyypillisimmän kuusijäsenisen klusterin mukaan laskettu klusteroitumisaste on rakenteen jäljessä:

AS-aineisto

hyvin <LA PRO VER *hyvin* ADJ LL>, 4,3
kovin <SUB VER VER *kovin* ADJ LL>, 5,1
oikein <?LA SUB VER *oikein* ADJ SUB>, 5,0

KSM-aineisto

hyvin <LA PRO VER *hyvin* ADJ LL>, 1,8
kovin <SUB VER VER *kovin* ADJLL>, 3,4
oikein <LA PRO VER *oikein* ADJ SUB>, 4,5

KSE-aineisto

hyvin <LA PRO VER *hyvin* ADJ LL>, 5,0
kovin <SUB VER VER *kovin* ADJ LL>, 4,2
oikein <PRO VER PRO *oikein* ADV LL>, 3,2 tai <LA PRO VER *oikein* ADJ LL>, 3,2

Kvalitatiivisesti tarkasteltuna *hyvin* ja *kovin* säilyttävät AS-aineistossa havaitun klusterirakenne-eron myös käänösaineistoissa, koska niiden tyypillisimmät kuusisanaiset kolligaattiklusterit ovat samoja kaikissa aineistoissa ja siten myös erot määritteiden välillä säilyvät. Sen sijaan *oikein*-määritteen klusteroituminen näyttää lähentävän määritettä *hyvin*-lekseemiin käänöskielessä: tämä tulee esiin KSM-aineistossa vasemmanpuoleisen kotekstin samankaltaisuudessa (*oikein*-määritteen klusterin vasemmanpuoleinen osa <LA PRO VER> on rakenteellisesti sama kuin *hyvin*-määritteelläkin) ja KSE-aineistossa kahdesta tyypillisimmästä klusterista toisen identtisessä rakenteessa *hyvin*-määritteen tyypillisimmän klusterin kanssa. Herkkyys klusteroitumiseen ei erottele määritettä alkuperäissuomessa eikä englannista käännettyssä suomessa, mutta KSM-aineistossa *hyvin*- ja *oikein*-määritteet eroavat merkittävästi klusteroitumisasteeltaan¹⁰³. Kvalitatiiviset ja kvantitatiiviset erot näyttävät siis ohjaavan jälleen hienokseltaan erilaiseen tulkintaan: klustereiden laatu osoittaa jonkinasteista samankaltaistumista, klusteroitumisherkkyys taas vähäistä taipumusta synonymian heikentymiseen. Kumpaan ilmiötä tukevat havaintojen määrät ovat kuitenkin suhteellisen vähäisiä, joten tulosta on pidettävä enemmänkin suuntaa antavana.

7.3.3. Paradigmaattinen morfologinen preferenssi

Astemääritteiden morfologisesta preferenssistä on käsitelty sekä alkuperäis- että käänössuomen analyyseissa paradigmaattista preferenssiä eli liitepartikkelien liittymistä astemääritteisiin. Alkuperäissuomea analysoitaessa *oikein* osoittautui määritteeksi, joka ei saa minkäänlaisia liitepartikkeleita suffikseikseen. *hyvin* puolestaan sai pelkästään *kin*-liitteitä, mutta *kovin* sekä *kaan*- että *kin*-liitteen, joskin jälkim-

¹⁰³ KSM: *hyvin* vs. *oikein* $z = 1,981372$ ($p < 0,05$).

mäisen osuus oli vähäinen. Määritteillä on siten alkuperäissuomessa hyvin selvä omanlaisensa paradigmaattinen morfologinen preferenssi. Alkuperäissuomen ja käännessuomen aineistojen liitepartikkelillisten muotojen osuudet kaikista astemääritteiden esiintymistä on lueteltu taulukossa 7.30.

Taulukko 7.30. Astemääritteiden paradigmaattinen preferenssi aineistoittain, liitepartikkelillisten muotojen prosentiosuudet kaikista esiintymistä.

| | | <i>hyvin</i> | <i>kovin</i> |
|---------------------|-------------|--------------|--------------|
| AS-aineisto | <i>kin</i> | 7,7 % | 1,7 % |
| | <i>kaan</i> | – | 8,0 % |
| KSM-aineisto | <i>kin</i> | 4,2 % | 2,9 % |
| | <i>kaan</i> | – | 15,9 % |
| KSE-aineisto | <i>kin</i> | 0,7 % | 0,3 % |
| | <i>kaan</i> | – | 7,8 % |

Luvut osoittavat ensinnäkin, että *hyvin*-määritteen *kin*-liitteisten muotojen osuus on pienempi KSM-aineistossa kuin alkuperäissuomessa ja lähes olematon englannista suomennetuissa teksteissä. Koska *kin*-liitteisiä muotoja voidaan pitää nimenomaan *hyvin*-määritteen tunnusmerkkinä piirteinä, vähenee tämän määritteen morfologinen tunnusmerkkinäisyys käännesskielessä. *kovin*-määritteellä tilanne on toisenlainen: sen selvästi tunnusmerkkinen *kaan*-liitteisten muotojen määrä on käännesskielessä kaksinkertainen, mutta lähdekieli näyttää kuitenkin vaikuttavan, sillä KSE-aineistossa osuus on likimain sama kuin AS-aineistossakin. Tämän määritteen tunnusmerkkinen *kaan*-liitteisten muotojen määrä näyttää siis yleensä lisääntyvän käännesskielessä. Määritteiden tunnusmerkkinen morfologinen preferenssi antaa siis vastakkaisen tuloksen määritteiden kesken: *hyvin*-määritteen tunnusmerkkinäisyys häviää, *kovin*-määritteen lisääntyy.

Jos määritteiden taipumusta paradigmaattiseen morfologiseen preferenssiin tarkastellaan kokonaisuutena, käy ilmi, että alkuperäissuomessa molemmilla määritteillä on lähes yhtä suuri taipumus saada liitepartikkeli. Liitepartikkelillisten muotojen osuus on *hyvin*-määritteellä 7,7 % ja *kovin*-määritteellä 9,7 %. Käännesskielessä tämä yhdenmukaisuus kuitenkin häviää, ja määritteiden morfologinen preferenssi on eriasteista: KSM-aineistossa osuudet ovat *hyvin*-määritteellä 4,2 % mutta *kovin*-määritteellä peräti 18,8 %, KSE-aineistossa osuudet ovat 0,7 % ja 8,1 %. Tämäkin korostaa sitä, että *kovin*-määritteen morfologinen preferenssi voimistuu käännettäessä, mutta *hyvin*-määritteen vähenee. *kovin* näyttää siten kasvattavan eroa synonyymeihinsa, mutta *hyvin* ja *oikein* jopa mahdollisesti lähenevät toisiaan käännesskielessä, koska toisen määritteen preferenssi osoittaa merkkejä laimentumisesta. Tulokset tuovat jälleen esiin määritekohtaiset erot ja lähdekielen vaikutuksen.

7.3.4. Semanttinen preferenssi ja prosodia

Miten sitten kontekstin semanttiset ominaisuudet erottelevat synonyymeja käännskielessä? Tätä käsittelyä varten on seuraavista kahdesta taulukosta ensimmäiseen koottu semanttisista preferensseistä ne merkitysluokat, joiden osuuksien erot ovat tilastollisesti merkitsevät ja erottelevat synonyymit toisistaan. Jälkimmäiseen taulukkoon on vuorostaan kerätty aineistoittain määritteiden semanttiset prosodiat ja niiden väliset erot.

Ensin tarkasteltava semanttinen preferenssi antaa astemääritteiden aineistokohdaisesta synonyymisyydestä taulukossa 7.31 esitetyn kaltaisen kuvan. Taulukkoon on koottu kaikki ne merkitysluokat, joiden osuudet poikkeavat astemääritteiden välillä tilastollisesti merkitsevästi, merkitsevyysarvo on myös ilmoitettu taulukossa. Alkuperäissuomen aineistossa on sellaisia merkitysluokkia, joiden osuuksien erot ovat merkitseviä ja siten erottavat astemääritteitä toisistaan, astemäärittepareittain neljä. Kuten huomata saattaa, käännskaineistoissa erottavien merkitysluokkien määrä on suurempi: lukumäärä vaihtelee viidestä seitsemään. Tämä on tulkittavissa siten, että käännskusuomessa on lähdekielestä riippumatta enemmän astemääritteitä erottavia merkitysluokkia O1-positiossa kuin alkuperäissuomessa ja siten synonyymisten astemääritteiden semanttiset preferenssit eroavat enemmän käännsk- kuin alkuperäissuomessa. Erot eivät siis kavennu käännskielessä vaan päinvastoin vahvistuvat. Tätä korostaa myös tilastollista merkitsevyyttä ilmoittavien tunnuslukujen joukossa erittäin merkitsevää eroa ilmoittavien tunnuslukujen ($z > z_{.001} = 3,29$) määrä: AS-aineistossa näin merkitsevä ero havaitaan neljän tarkastellun merkityspiirteen kohdalta, KSM-aineistossa kuuden ja englannista käännketyissä teksteissä seitsemän merkityspiirteen kohdalta. Koska tilastollinen merkitsevyys on yksinkertaistetusti sitä suurempi, mitä suurempi on laskettavien piirteiden osuuksien välinen ero, voidaan myös näiden tunnuslukujen perusteella päätellä synonymian heikkenevän määritteiden välillä käännskusuomessa.

Semanttisen prosodian negatiiviset ja positiiviset valinnat jakautuivat alkuperäissuomen aineistossa vuorostaan siten, että *kovin*-määritteellä oli selkeä negatiivinen ja *oikein*-määritteellä positiivinen prosodia, mutta *hyvin*-määritteen valinta ei ollut yhtä selkeä vaan jakautui liki puoliksi molempiin. Prosodia olikin selvästi synonyymit toisistaan erottava kontekstuaalinen piirre. Taulukkoon 7.32 on koottu negatiivisten ja positiivisten prosodioiden osuudet astemääritteittäin ja aineistoittain. Vertailussa synonymian muuttumista tarkastellaan prosenttimäärien erotuksen muuttumisen avulla.

Taulukko 7.31. Astemääritteitä erottelevat semanttiset preferenssit aineistoittain.

| | <i>hyvin vs. kovin</i> | <i>hyvin vs. oikein</i> | <i>kovin vs. oikein</i> |
|-----------------|------------------------|-------------------------|-------------------------|
| AS-aineisto | | | |
| määrä | z=3,87 (p<0,001) | arvo | z=10,28 (p<0,001) |
| arvo | z=3,22 (p<0,01) | aika/ikä | z=4,20 (p<0,001) |
| vaikeus | z=2,78 (p<0,01) | tuttuus | z=3,14 (p<0,01) |
| f tila/aistimus | z=2,73 (p<0,01) | f tila/aistimus | z=2,57 (p<0,05) |
| KSM-aineisto | | | |
| määrä | z=5,62 (p<0,001) | arvo | z=12,20 (p<0,001) |
| ulkonäkö | z=3,06 (p<0,01) | aika/ikä | z=3,77 (p<0,001) |
| intensiteetti | z=2,66 (p<0,01) | ulottuvuus | z=3,05 (p<0,01) |
| arvo | z=2,42 (p<0,05) | tuttuus | z=2,52 (p<0,05) |
| varmuus | z=2,41 (p<0,05) | paikka | z=2,05 (p<0,05) |
| | | vaikeus | z=2,03 (p<0,05) |
| KSE-aineisto | | | |
| arvo | z=5,82 (p<0,001) | arvo | z=10,37 (p<0,001) |
| määrä | z=3,76 (p<0,001) | aika/ikä | z=2,93 (p<0,01) |
| f tila/aistimus | z=3,73 (p<0,001) | tuttuus | z=2,80 (p<0,01) |
| ulkonäkö | z=3,66 (p<0,001) | tunne | z=2,60 (p<0,01) |
| paikka | z=2,64 (p<0,01) | ulottuvuus | z=2,50 (p<0,05) |
| tunne | z=2,17 (p<0,05) | vaikeus | z=2,28 (p<0,05) |
| | | f tila/aistimus | z=2,06 (p<0,05) |

Taulukko 7.32. Astemääritteiden semanttinen prosodia aineistoittain.

| | | <i>hyvin</i> | <i>kovin</i> | <i>oikein</i> |
|--------------|--------------|--------------|--------------|---------------|
| AS-aineisto | positiivinen | 42,3 % | 8,5 % | 70,2 % |
| | negatiivinen | 50,0 % | 89,8 % | 28,1 % |
| | neutraali | 7,7 % | 1,7 % | 1,7 % |
| | yhteensä | 100 % | 100 % | 100 % |
| KSM-aineisto | positiivinen | 48,8 % | 9,2 % | 78,3 % |
| | negatiivinen | 39,7 % | 88,4 % | 19,7 % |
| | neutraali | 11,6 % | 2,4 % | 1,9 % |
| | yhteensä | 100 % | 100 % | 100 % |
| KSE-aineisto | positiivinen | 38,2 % | 8,5 % | 78,8 % |
| | negatiivinen | 49,7 % | 90,5 % | 18,6 % |
| | neutraali | 12,1 % | 1,0 % | 2,6 % |
| | yhteensä | 100 % | 100 % | 100 % |

Astemääritteistä *kovin*-lekseimin prosodia on aineistoittain samanlainen, sillä sen tunnusmerkkisen negatiivisen prosodian osuus on samansuuruinen, noin 90 % kaikissa aineistoissa. Sen sijaan *hyvin*-määritteen prosodiset piirteet muuttuvat käännessuomessa hieman: positiivisen ja jossain määrin myös neutraalin prosodian osuudet kasvavat ja negatiivisen osuus pienenee KSM-aineistossa. Tässä mielessä

määrite näyttää käännösuomessa erottuvan enemmän *kovin*-määritteestä, sillä *ko-**vin*-määritteelle tunnusmerkkisen negatiivisen prosodian osuus on *hyvin*-määritteellä vähäisempi, mutta samanaikaisesti se hieman lähenee *oikein*-määritettä. Tämä ei koske kuitenkaan englannista suomennettuja tekstejä. Neutraalin prosodian lievää lisääntymistä voi pitää myös merkinä jonkinasteisesta prosodian neutraalistumisesta, eli määritteen jo alkuperäissuomessa tunnusmerkitön prosodia näyttää korostuvan hieman käännöskielessä. *oikein*-määritteen positiivisen prosodian osuus näyttää vuorostaan kasvavan entisestään käännösuomessa. Muutos näkyy negatiivisen prosodian pienenemisenä, mutta ei neutraalisuuden lisääntymisenä, joten määritteelle tunnusmerkkinen prosodin piirre näyttää korostuvan käännösuomessa entisestään. Prosodia-analyysin perusteella *oikein*-määritteen synonymiaa rikkova prosodin ominaisuus näyttää erkaannuttavan sitä synonyymeistaan käännösuomessa; *hyvin*-määritteen prosodiset piirteet eivät yksiselitteisesti lähennä sitä synonyymeihinsa tai erkaannuta sitä synonyymeistaan, pikemminkin vertailu osoittaa aineiston merkitystä muutoksessa ja mahdollisesti lievää piirteiden neutraalistumista.

7.4. Kokoava katsaus: oletuksia ja havaintoja käännöskielen leksikaalisista ominaisuuksista

Edellä olevassa analyysissä tuli esiin lukuisia yksityiskohtaisia seikkoja, jotka kuvaavat käännösuomen lekseemien kontekstuaalista valintaa ja synonymiaa. Miten nämä tulokset sitten vastaavat luvussa 5 esitettyihin käännöskieltä koskeviin hypoteeseihin ja suhteutuvat aiempaan tutkimukseen? Tätä tarkastellaan seuraavaksi olettamus kerrallaan.

Epätyypilliset frekvenssit. Oletus käännösuomessa tavattavista epätyypillisistä esiintymistäajuuksista koski sekä lekseemien että kontekstuaalisten piirteiden esiintymistä. Alkuperäissuomelle epätyypillinen frekvenssi voi toteutua käännöksissä joko kieliaineksen ali- tai yliedustumisena; epätyypillisiä esiintymistäajuuksia testattiin tilastollisen merkitsevyyden avulla. Aineistosta oli löydettävissä selvää tukea oletukselle ja aiemmille havainnoille käännöskielestä tavattavista epätyypillisistä frekvensseistä. Ensinnäkin kaikki astemääritteet olivat vähintäänkin lievästi yliedustuneita käännöskielessä, mikä tuli ilmi Johanssonin (1980) käyttämää laskentaa mukailevassa sanafrekvenssitestissä. Kokonaisuutena kaikkien kolmen määritteen yliedustumista oli havaittavissa khin neliö -testin perusteella vain KSM-aineistossa, mutta normitettujen frekvenssien ja edellä mainitun testin tulosten yhdenmukaisuus KSM- ja KSE-aineistojen kesken todistaa kuitenkin, että määritteet yliedustuvat molemmassa käännösaineistoissa. Tutkimuksessa tarkasteltuja astemääritteitä käytetään siis käännösuomessa enemmän kuin alkuperäissuomessa. Myös monien kontekstuaalisten piirteiden osuus oli käännösuomessa joko selvästi pienempi tai suurempi kuin alkuperäissuomessa. Ali- ja yliedustumista oli havaittavissa kaikissa kontekstuaalisissa merkityspiirteissä aina yksittäisten kollokaattien epätyypillisistä

frekvensseistä prosodisten piirteiden epätyypillisiin osuuksiin. Epätyypilliset frekvenssit näyttivät kuitenkin kasautuvan tiettyihin kontekstuaalisiin ominaisuuksiin. Erityisesti kollokaattien epätyypilliset esiintymistäajuudet (mikä on perusteena epätyypillisille kollokoinnille) ja yleisimpien kollokaattien osuudet osoittautuivat alkuperäissuomelle epätyypilliseksi. Kolligatiivinen piirre, jossa ilmeni runsaasti alkuperäissuomelle epätyypillisiä frekvenssejä, oli sanaluokkakolligaattien esiintyminen. Sen sijaan morfologiset ja semanttiset piirteet eivät olleet yhtä herkkiä alitai yliedustumaan käännskielellä, vaikkakin epätyypillisiä määriä tavattiin näistäkin piirteistä. Erityisesti *hyvin*-määritteen kontekstista löytyi epätyypillisiä frekvenssejä myös viimeksi mainituissa piirteissä.

Epätyypillisiin frekvensseihin liittyvä tutkimustulos antaa olettaa, että leksikaalisia myötäesiintymiä abstraktimmat kontekstuaaliset piirteet eivät vaihtelee suuresti, kun tarkastellaan yhden kielen ja tekstilajin käänös- ja alkuperäismuotoja. Erityisesti kontekstin semanttisiin piirteisiin liittyvät ominaisuudet, semanttinen preferenssi ja prosodia, jotka muodostavat kontekstuaalisen merkitysyksikön uloimmat kerrokset, ovat suhteellisen jähmeitä muuttumaan. Semanttisen preferenssin piirre-analyysi osoitti muutoksia vain harvojen piirteiden kohdalla, ja myös negatiivisen ja positiivisen prosodian osuudet olivat lähes samanlaisia aineistoittain – tässä kuitenkin *hyvin*-määritteen kontekstuaaliset piirteet poikkesivat sen synonyymeista. Vaikuttaakin siltä, että semanttiset ja oletettavasti leksikaalisten yksiköiden valintaa ohjaavat prosodiset ominaisuudet ovat suhteellisen jähmeitä muuttumaan ja siten säilyvät myös käännskielellä samanlaisina kuin alkuperäiskielelläkin.

Epätyypilliset leksikaaliset kombinaatiot. Tutkimuksen toinen päähypoteesi oletti, että käänössuomesta on löydettävissä sellaisia kontekstuaalisia merkityspiirteitä, jotka ovat alkuperäissuomelle epätyypillisiä. Tätä oletusta tarkasteltiin tutkimalla *hyvin*-, *kovin*- ja *oikein*-määritteen kontekstuaalista merkitystä kolmessa eri aineistossa. Analyysi toi esille epätyypillisiä kombinaatioita astemääritteen kollokatiivisissa merkityksissä, sillä käänössuomen kollokaatioista löytyi sellaisia myötäesiintymiä, joita ei tavata tyypillisinä alkuperäissuomessa. Tyypillisuus määriteltiin analyyseissa tilastollisen merkitsevyyden perusteella. Tulos on yhdenmukainen Maurasen (2000b) aiempien käänössuomen epätyypillisiä kollokaatiivisia suhteita tarkastelleen analyysin kanssa, jossa kollokaatiivisten suhteiden muuttumista käsiteltiin muun muassa *toisaalta*- ja HALUTA- ja PYRKIÄ-lekseemien avulla.

Aineistosta oli löydettävissä sekä sellaisia leksikaalisia myötäesiintymiä, joita tavattiin molemmista käänössuomen korpuksista (mutta ei alkuperäissuomen), ja myös sellaisia, jotka olivat tyypillisiä vain jommallekummalle käänösaineistolle. Ensin mainittuja olivat muun muassa kollokaatiot *hyvin* <YLPEÄ>, *kovin* <KUULOSTAA> ja *oikein* <MUKAVA>. Pelkästään englannista käännettyissä teksteissä havaittiin puolestaan sellaisia myötäesiintymiä kuin *hyvin* <NÄYTTÄÄ>, *kovin* <SURULLINEN> ja *oikein* <KIITOS>. Myös käännskielelle vieraita myötäesiintymiä oli löydettävissä: vain alkuperäissuomelle tyypilliseksi osoittautuivat kollokaatiot *hyvin* <*hiljaa*> ja *oikein* <*tuntua*>. Viimeksi mainitut esiintymät vahvistavat tuloksia käännskielen osittain erilaisesta leksikaalisesta luonteesta, sillä jos käänös-

kielessä on sellaisia kollokaatioita, joita ei tavata alkuperäissuomessa, näyttää siitä myös puuttuvan myötäesiintymiä, jotka ovat alkuperäissuomelle tyypillisiä.

Myös klusterirakenteista oli löydettävissä alkuperäissuomelle epätyypillisiä leksikaalisia ketjuja. Erityisesti tämä tuli esiin kolmisanaisia klustereita pitemmissä rakenteissa. Tarkastelussa muut mukana olleet kontekstuaaliset ominaisuudet – kolligationaalinen valinta, morfologinen preferenssi sekä semanttinen preferenssi ja prosodia – osoittautuivat kollokationaalista valintaa jähmeämmiksi muuntumaan. Niistä ei juuri tavattu epätyypillisiä kombinaatioita, poikkeuksena mahdollisesti *oikein*-määritteen kolligaattiklusterit. Näiden analyysissa aineisto tuotti kuitenkin niin vähän esiintymiä, että tulosta on pidettävä pikemminkin suuntaa antavana.

Käännöskielen synonymia. Ensimmäinen synonyymisten suhteiden muutosta koskeva analyysi käsitteli synonyymien keskinäisiä suhteita, eli sitä, onko jokin käännösten synonyymisista määritteistä selvästi yleisempi kuin toinen määrite. Tämä ei saanut tukea aineistosta: astemääritteiden osuudet olivat aineistoittain erittäin samanlaiset ja kaikkein frekventeimmän määritteen (*hyvin*) osuus oli tismalleen sama kaikissa aineistoissa. Tulos ei siten tue aiempia oletuksia (Blum-Kulka ja Levenston 1983; Klaudy 1996; Kohn 1996) synonyymian yksinkertaistumisesta vaan päinvastoin vahvistaa omaa aikaisempaa astemääritteillä tehdyn tutkimukseni tulosta (Jantunen 2001a), jossa ei myöskään löytynyt evidenssiä synonyymipesyeen tyypillisimpien jäsenten yliedustumiselle käännöskielessä. Tukea ei saatu myöskään Maurasen (2000b) havainnolle, jonka mukaan käännöksissä suositaan synonyymipesyeestä toisia ilmauksia kuin alkuperäiskielisissä teksteissä.

Jotta synonyymipesyeen kirjo voisi ylipäättään kaventua käännöskielessä, olisi synonyymisten ilmausten lähennyttävä toisiaan kontekstuaalisesti niin, että ne olisivat korvattavissa toisillaan. Kontekstuaalista lähentymistä olisi tapahduttava erityisesti sellaisten ilmausten kesken, joilla on hyvin selkeät kontekstuaaliset valintapiirteet. Luultavaa kuitenkin on, että vaatimus kontekstien samankaltaistumisesta koskisi synonymiaa yleisestikin, koska kontekstuaalinen merkitys on selvästi synonyymeja erottava tekijä. Jos kontekstuaaliset ominaisuudet yhdenmukaistuisivat joidenkin ilmausten välillä, voisi jokin synonyymipesyeen jäsenistä syrjäyttää toisen ilmauksen, esimerkiksi frekventti harvaan esiintyvän, ja näin synonyymisten ilmausten kirjo voisi ainakin teoriassa kaventua kielessä.

Ovatko käännöskielen synonyymiset astemääritteet sitten kontekstuaalisesti enemmän toistensa kaltaisia eli onko synonyymisyyden aste voimistunut, vai ovatko ne vaihtoehtoisesti erkaantuneet kontekstuaalisesti toisistaan, jolloin määritekohtaiset ominaisuudet ovat korostuneet ja määritteen korvaaminen toisella olisi ainakin kontekstuaaliselta pohjalta vaikeampaa kuin alkuperäiskielessä? Tätä kysymystä käsittelevässä luvussa 7.3 selvisi, että yhteisten tilastollisesti merkitsevien kollokaattien määrän ja kaikille yhteisten kollokaattien määrän perusteella synonyymit lähenevät suomennoksissa hieman toisiaan. Tämä koskee etenkin astemääritepareja *hyvin–kovin* ja *kovin–oikein*. Koko tarkastelukotekstin sanastollista variatiota kuvaavien LSS-arvojen vertailu osoittaa kuitenkin päinvastaista: erot kasvavat selvästi käännösaineistossa, mikä tarkoittaa käytännössä sitä, että *hyvin*-määritteellä on laaja kollokationaalinen diversiteetti käännössuomessa ja sen synonyymeilla selvästi suppea-alaisempi kollokaattikirjo. LSS-arvot eivät kuitenkaan pal-

jasta kollokaattien laatua vaan ilmaisevat ainoastaan erilaisten kollokaattien suhdearvon. Arvot eivät siis paljasta suoranaisesti määritteiden korvattavuutta synonyymilla, vaan kertovat yhden synonyymisten sanojen kontekstuaalisen ominaisuuden muutoksesta käännskielessä.

Kollokaattiklustereissa kvalitatiivinen ero säilyy, mutta taipumus klusteroitumiseen tasoittuu käännskielessä. Määritteiden semanttisen preferenssin ja prosodian tarkastelu vuorostaan osoitti, että käännskielessä synonyymiset määritteet pikeminkin eroavat kontekstuaalisilta ominaisuuksiltaan selvemmin toisistaan kuin lähenisivät toisiaan. Tämän puolesta puhuivat useiden merkitysominaisuuksien suuremmat erot määritteiden välillä käännsuomessa. Synonyymien kontekstin lekseemit saattavat siis lähentyä toisiaan käännskielessä, mutta tämä vaatisi kuitenkin perusteellisempaa koko kontekstuaalisen variaation analysointia määritepareittain. Abstraktisempien valintapiirteiden perusteella synonyymit taas näyttävät eroavan käännsuomessa enemmän toisistaan.

Leksikaalinen yksinkertaistuminen. Luvun 7 analyseissa on viisi mittaria, joilla voidaan arvioida astemääritteiden kontekstuaalisten piirteiden yksinkertaistumista. Kollokaattien ja kolligaattien klusteroitumisasteita sekä kontekstilauseiden pituutta tarkastelemalla voidaan selvittää kontekstin rakenteellista yksinkertaistumista. Kontekstin frekventimpien kollokaattien osuudet ja leksikaalista diversiteettiä kuvaavat arvot kertovat puolestaan sanastollisesta yksinkertaistumisesta.

Kontekstin leksikaalisesta yksinkertaistumisesta käännskielessä saatiin selviä todisteita vain *oikein*-määritteen kontekstista, jossa leksikaalinen diversiteetti osoitautui käännskielessä pienemmäksi; lähdekieli ei vaikuttanut suuntaukseen. Frekventimpien kollokaattien osuudet olivat käännsuomessa yleensä yhtä suuret kuin alkuperäiskielelläkin eli näiltä osin yksinkertaistumista ei ollut selvästi havaittavissa. Tässä tulokset ovat erilaisia kuin Laviosan (Laviosa-Braithwaite 1996) ja Nevalaisen (tulossa) havainnot – on kuitenkin korostettava, että tutkimukset eivät ole täysin verrannollisia, koska jälkimmäiset tarkastelivat frekventtien sanojen osuutta koko aineistossa ja omassa tutkimuksessani keskityttiin vain yksittäisten lekseemien kontekstuaaliseen variaatioon. Vaikka frekventit kollokaatit eivät yleensä olleet taajemmin esiintyviä käännskielessä, teki tähän poikkeuksen kuitenkin englannista käännettyjen tekstien aineisto, jossa *kovin*- ja *oikein*-määritteiden kontekstien taajaan esiintyvät kollokaatit olivat muita aineistoja frekventtimmin käytettyjä. Kontekstien leksikaalista yksinkertaistumista ei tämän tuloksen valossa voida pitää yleisesti tutkittuja astemääritteitä koskevana, vaan yksinkertaistuminen voi olla sekä määrite- että aineistokohtaista.

Kontekstuaalisten rakennepiirteiden yksinkertaistumista oli havaittavissa jossain määrin. *hyvin*- ja *oikein*-määritteen leksikaalinen klusteroituminen oli voimakkaampaa käännskielessä, mutta *kovin*-määritteen kuitenkaan ei. Tyypillisimpien kolligaattirakenteiden osuus ei sen sijaan ollut suomennoksissa suurempi yhdelläkään määritteellä. Taipumusta voimakkaampaan klusteroitumiseen ilmeni siis vain leksikaalisella tasolla. Viitteitä käännskielen voimakkaammasta klusteroitumisesta on havaittavissa myös Kemppasen (tulossa) ja Tirkkonen-Conditin (tulossa) tutkimuksissa. Näiden tutkimusten valossa näyttääkin siltä, että käännskieli sisältää

enemmän muuttumattomia leksikaalisia ketjuja kuin alkuperäissuomi, ja on siten useampisanaisissa leksikaalisissa ketjuissa variaatioiltaan yksinkertaisempaa.

Kotekstuaalinen lausepituus ei osoittautunut käännössuomessa yleensä lyhyemmäksi kuin alkuperäissuomessa, joten tulos tukee Laviosan (Laviosa-Braithwaite 1996) kertomakirjallisuutta koskevia havaintoja; tosin Laviosan tutkimuksessa lausepituus oli jälleen laskettu koko aineistosta. Oma tutkimukseni paljastaa kuitenkin selvästi, että lähdekieli voi vaikuttaa kotekstilausepituuteen, sillä lausepituus osoittautui järjestelmällisesti kaikilla määritteillä lyhyemmäksi englannista suomennetuissa teksteissä.

Oletus siitä, että käännöskielen kotekstuaaliset profiilit eivät yksinkertaistu käännöskieleessä, osoittautui tutkimuksen perusteella suurelta osin oikeaksi, kun astemääritteitä ja aineistoa tarkastellaan kokonaisuutena. Astemääritekohtaisesti on kuitenkin havaittavissa merkkejä kotekstuaalisten piirteiden yksinkertaistumisesta.

Yhden lähdekielen vaikutus. Aineisto oli rakennettu siten, että yhden suomennoksia sisältävän tekstimateriaalin lähdekielenä oli pelkästään englanti. Tämän tarkoituksena oli luoda asetelma, joka mahdollistaa yhdestä lähdekielestä käännetyn aineiston vertailun käännöskielen aineistoon, jossa mikään lähdekieli ei ole muita edustuneempi. Läpi luvussa 7 esitetyn käännössuomen analyysin tulee voimakkaasti ilmi se, että yhdestä lähdekielestä, tutkimuksessani englannista, käännetty suomen kieli voi olla ominaisuuksiltaan erilaista kuin suomennosten suomi ”keskimäärin”, eli oletus lähdekielen vaikutuksesta vahvistuu. Englannista käännetty teksti poikkesivat monilähdekielisestä aineistosta selvästi kollokationaalisissa valinnoissa (sekä tilastollisten kollokaattien kvalitatiivisissa että kvantitatiivisissa ominaisuuksissa ja klusteroitumisessa), kolligationaalisissa preferensseissä – erityisesti lausepituuden suhteen –, morfologisessa preferenssissä ja semanttisten piirteiden osuussuhteissa jossain määrin. Suurimmillaan kahden käännösaineiston erot olivat *hyvin*-määritteen kontekstissa, sen synonyymeilla erot näyttivät olevan vähäisempiä. Synonymian muuttumista koskevassa analyysissä käännösaineistot sen sijaan näyttivät osoittavan yleisesti ottaen yhteneväistä linjaa: mikäli määritteen samankaltaisuus lisääntyi tai heikkeni monilähdekielisessä aineistossa, tapahtui vastaava ilmiö tavalisesti myös englannista käännettyissä teksteissä. Tässä havaitut tulokset tukevat voimakkaasti tähänastisia havaintoja käännösten lähdekielen vaikutuksesta käännöskieleen (erityisesti Eskola 2002; ks. myös Laviosa-Braithwaite 1996; Mauranen 2000b).

8. LOPPUKATSAUS

Työlläni oli kaksi tavoitetta, mikä näkyy myös esityksen rakenteessa. Ensimmäiseksi tehtäväkseni otin tarkastella, miten synonyymiset ilmaukset eroavat toisistaan lähimmän tekstiyhteyden perusteella. Tutkimuksessa pyrittiin luomaan synonyymeille kontekstuaalinen profiili, joka kuvaa niiden kontekstissa – pääasiassa kontekstissa – ilmeneviä preferenssejä. Oletus synonyymisten ilmausten kontekstisidoksisuudesta juontui lekseemeille tyypillisestä kontekstuaalisesta valinnasta eli kontekstuaalisesta merkityksestä. Joissakin aiemmissa tutkimuksissa (Persson 1990; Jantunen 2001b; Arppe 2001, 2002) synonyymisten lekseemien kontekstuaalisia preferenssejä onkin analysoitu, mutta systemaattinen tarkastelu nimenomaan ko(n)tekstilähtöisestä ja ko(n)tekstin erilaiset preferenssit läpäisevistä ja ne yhdistävistä piirteistä on vielä ollut olematonta. Tutkimukseni tavoitteena oli esittää yksi tällainen analyysi ja tarkastella esimerkin kautta erityisesti sitä, miten synonyymiset sanat *eroavat* kontekstuaaliselta valinnaltaan toisistaan.

Toisena tavoitteenani oli vertailla käännössuomen sanaston kontekstuaalisia piirteitä alkuperäissuomen ominaisuuksiin eli testata oletusta käännössuomen epätyypillisistä leksikaalisista kombinaatioista. Epätyypillisten kombinaatioiden tutkimus nimenomaan verrannollisilla korpuksilla tehtynä on ollut hyvin vähäistä, ja varsinainen tutkimuskysymyksen on herännyt vastikään (Mauranen 2000b). Myös tämä tutkimus kaipasi laajentamista kontekstin läpäiseviin moninaiisiin piirteisiin, sillä aiempi tutkimus on ottanut huomioon pääasiassa vain leksikaaliset myötäesiintymät. Epätyypillisten kombinaatioiden tarkastelun ohessa työssä pyrittiin tuomaan valoa jo aiemmin tarkemminkin käsiteltyihin käännöskielen ominaisuuksiin, kuten kielen yksinkertaistumiseen ja lähdekielen vaikutukseen. Lisäksi haluttiin myös tutkia, mitä synonymialle tapahtuu käännettäessä, aiemmin on oletettu, että synonyymiakirjo kaventuisi käännettäessä kielestä toiseen ja olisi myös erilainen yhden kielen käännös- ja alkuperäismuotojen välillä (Blum-Kulka ja Levenston 1983; Klaudy 1996; Kohn 1996; Mauranen 2000b).

Päämääriin pyrin kolmen synonyymisen astemääritteen eli lekseemien *hyvin*, *koivin* ja *oikein* avulla. Astemääritteet ovat monella tapaa mielenkiintoisia tutkimuskohteita työn tavoitteiden kannalta: ne ovat suhteellisen yleisiä kielenaineiksia, niiden eri ryhmät sisältävät runsaasti synonyymiaa eikä niiden käyttö ole tekstin aihepiirin motivoimaa. Lisäksi ne ovat lähellä kieliopillisia sanoja, joiden mahdollisuudesta synonyymisyyteen on esitetty vastakkaisia näkemyksiä (ks. Kangasniemi 1997; Kuiri 2000). Sekä synonyymisyyteen liittyvistä kontekstuaalisista ominaisuuksista että käännössuomen kuvaukseen liittyvistä tuloksista on esitetty kokoavat katsaukset kyseisiä teemoja käsittelevien lukujen lopussa, joten en kertaakaan niitä enää tässä. Sen sijaan on vielä paikallaan pohtia kysymyksiä, joita työ nosti esiin ja jotka vielä kaipaisivat vastauksia.

Tutkimuksen perusasetelma eli keskittyminen kolmen synonyymisen ilmauksen analysoimiseen asettaa tulosten yleistettävyydelle luonnollisesti rajoituksia. Tällä tavalla menettelemällä saatiin kattavasti tietoa niistä pääasiassa kontekstin, mutta paikoin myös laajemman kontekstin ominaisuuksista, jotka voivat erottaa hyvinkin

lähimerkityksiset synonyymit toisistaan. Työssä esitettyä noodin ja sen lähiympäristön kuvausta onkin pidettävä pikemminkin menetelmällisenä pyrkimyksenä, sillä tuloksia ei voi yleistää laajasti leksikkoa koskevaksi; jo pelkästään astemääritteidenkin ryhmä on niin heterogeeninen, että tutkimuksen kolmen määritteen välillä havaittujen erojen laajentaminen koskemaan muita astemääritteitä vaatisi tarkistuksia. Oletettavaa kuitenkin on, että samanlaisia eroja on löydettävissä. Miten laajasti eri ko(n)tekstuaalisilla tasoilla havaitut erot ja yhtäläisyydet koskevat muiden sanaluokkien ja esimerkiksi merkitykseltään konkreettisempia synonyymeja, vaatii vielä jatkotutkimusta. Erityisesti tämä koskee morfologista preferenssiä, jonka analyysiin astemääritteet eivät taipumattomina kovinkaan hyvin sovellu. Tutkimuksen voitaneenkin katsoa tarjoavan erilaisia työkaluja lähimerkityksisten sanojen kontekstuaalisten piirteiden analysointiin.

Paikoin tutkimustyölle luonteenomainen haasteellisuus muuttui vaikeudeksi. Osittain tähän johtivat valitut havaintoyksiköt, valittu tekstilaji ja mahdollisuuksien mukaan käytetty korpusvetoinen lähestymistapa. Havaintoyksiköiden eli astemääritteiden merkityksen abstraktisuus ja puolikieliopillinen luonne yhdistyvät kotekstin runsaaseen variaatioon, mikä näkyy etenkin semanttisen preferenssin analyysissa monina merkitysluokkina; luultavaa on, että merkitykseltään konkreettisemmat ja selkeärajaisemmat ilmaukset tuottaisivat vähemmän variaatiota sisältävän preferenssianalyysin, jolloin oltaisiin lähempänä sinclairiläistä yksiselitteistä semanttisen preferenssin analyysia. Toisaalta keskittyminen aikaisempaa tutkimusta abstraktimpiin ilmauksiin tuo esiin aiemmassa tutkimuskirjallisuudessa esiintyneen yksipuolisuuden ja yksinkertaistuksen: ilmausten semanttinen preferenssi ei olekaan aina kuvattavissa niin selkeänä ja tarkkarajaisena kuin on esitetty.

Merkitykseltään abstraktien havaintoyksiköiden ja niiden kotekstin semanttisten valintojen tarkastelu nostaa esille myös tutkijan oman intuition lisääntymisen muuten intuition määrän minimoimiseen pyrkivässä korpuslingvistiikassa. Kun kollokaatio-, kolligaatio- ja morfologiasuhteet ovat suhteellisen yksiselitteisiä ja olemassa oleviin luokitteluihin jälkimmäisten osalta perustuvia, ei intuitio saa niiden analyysissa kovin paljon sijaa. Sen sijaan merkityssuhteiden kanssa on tilanne toisenlainen: jo semanttisen preferenssin analyysia tehtäessä huomattiin, että pelkästään adjektiivien luokitteluun on olemassa monia – intuition perustuvia – ryhmityksiä, joiden soveltaminen preferenssianalyysin sanaluokkajaot ylittävän analyysin perusvaatimukseen osoittautui vaikeahkoksi. Intuition määrä lisääntyy entisestään kotekstin yksittäisten kollokaattien merkityspiirteitä analysoitaessa ja myös tekstiyhteydestä tai toisinaan jopa yleisestä ajattelutavasta haettavan semanttisen prosodian analyysissa. Näitä analysoidessaan tutkija on pitkälti oman intuitiionsa varassa, jolloin myös toisenlaisia tulkintoja voidaan esittää. Tulkinnan ongelmaan liittyy kiinteästi myös aineistona ollut tekstilaji, joka jo itsessään sisältää inhereentisti monitulkintaisuuden toisin kuin vaikkapa täsmällisiin merkityksiin pyrkivä akateeminen tekstilaji.

Synonymian kontekstuaalisuuden käsittely tuotti kuitenkin havaintoja, jotka – sen lisäksi, että ne osoittautuivat merkittäviksi nimenomaan käsiteltyjen astemääritteiden analyysissa – ovat mielenkiintoisia ja pohdinnan arvoisia ulottuvuuksia laajemminkin synonymian ja kontekstuaalisten piirteiden selvityksissä. Ensinnäkin suomen kielen kielto- ja myöntöhakuisten ilmausten kuvaukseen liittyen tutkimuk-

nessa tuli ilmi, että astemääritteistä osa voi kuulua selvästi jompaankumpaan ryhmään: tutkimus paljasti, että *kovin* on hyvin vahvasti kieltohakuinen määrite, *oikein* taas myöntöhakuinen. Vastaavanlaista distribuutiota lienee mahdollista löytää myös muista astemääriteryhmistä; Sadeniemi (1942: 286) mainitseekin sellaiset moderaatiiviset astemääritteet kuin *melko* (*suuri*), *melkein* (*täydellinen*) ja *jokseenkin* (*kallis*) vain myönteisissä yhteyksissä esiintyviksi. Aineiston analyysissä selvisi myös, että vaikka määritteellä ei olisikaan kieltohakuisuuden piirrettä eli preferenssiä kieltomuotoiseen kontekstiin (lauseeseen), voi konteksti paljastaa, että sitä käytetään merkittävässä määrin kuitenkin merkitykseltään kielteisessä kontekstissa. Tällaista tilannetta kuvaa mainiosti *hyvin*-määritteen analyysi: määritteen kontekstissa kieltomuotoisten lauseiden osuus on ainoastaan 0,6 %, mutta sen semanttinen prosodia osoittautuu puoliksi negatiivisen ja positiivisen prosodian kesken jakautuvaksi. Määritteen osittainen negatiivinen prosodia ei synnykään kieltoverbin läsnäolosta vaan muista merkitykseltään kielteisistä ilmauksista ja tilannekontekstista. Kieltomutoisiin rakenteisiin keskittyy kieltohakuisuuden analyysia voitaisiinkin täydentää kontekstin semanttisten ominaisuuksien analyysillä, jolloin ilmausten ”kielto- ja myöntöhakuisuudesta” syntyisi monipuolisempi kuva.

Tutkimuksessa tuli esiin myös piirre, joka on tähdellistä nostaa vielä käsittelyyn synonyymisten ilmausten kontekstuaalisen eroavuuden tutkimisen kannalta. Astemääritteiden kollokaatioanalyysissä selvisi, että sellaisia kollokaatteja, jotka yhdistivät näitä synonyymisiä astemääritteitä, oli vähiten noodien vieruspositioissa VI ja O1, ja mitä kauemmaksi noodista kuljettiin, sitä enemmän yhdistäviä tekijöitä löytyi. Tämän perusteella voitaisiin olettaa, että kullekin synonyymipesyteen jäsenelle tyypilliset leksikaaliset valinnat saattavat erottua jo lähipositioista ja vastaavasti kontekstin merkitys neutralisoituu sitä enemmän, mitä kauemmaksi noodista menään.

Tälle löytyy vahvistusta synonyymisten ilmausten tilastollisesti merkitsevien kollokaattien analyysistä: ne kollokaatit, jotka osoittautuivat merkitseviksi molempien työssä käytettyjen merkitsevyydestien perusteella, olivat yhtä poikkeusta lukuun ottamatta sellaisia kollokaatteja, joita tavataan pääasiassa oikeanpuoleisesta vieruspositiosta. Luonnollisesti synonyymeja erottelevia kollokaatteja tavataan kauempaakin kuin vain vieruspositioista, mutta lienee todennäköistä, että synonyymikohtaiset kollokationaaliset valinnat ovat havaittavissa jo aivan lähikotekstia, jopa vieruspositioita analysoitaessa. Tulosta voidaan tarkastella suhteessa Masonin (1997, 2000) esittämiin havaintoihin leksikaalisesta gravitaatiosta. Hänen havaintojensa mukaan leksikaalinen variaatio suurenee yleensä sitä mukaa, mitä kauemmaksi noodista kuljetaan, ja lähipositioissa variaatio on usein pienimmillään. Masonin tulos on tulkittavissa siten, että lähipositioiden leksikaalisen variaation vähäisyys johtuu pitkälti juuri kollokationaalista valinnasta, joka esittyytyy noodin lähikotekstissa.

Aiempi TÄRKEÄ- ja KESKEINEN-adjektiiveista tekemääni tutkimus (Jantunen 2001b) tukee myös havaintoa vieruspositioiden merkityksestä synonyymien kontekstuaalisten valintojen tutkimuksessa, sillä juuri vieruskollokaattien avulla pystyttiin osoittamaan näiden adjektiivien kollokatiiviset valinnat perin erilaisiksi. Tässä valossa Vanhatalon (2003) samoihin adjektiiveihin keskittyneessä kyselytestitutkimuksessa esittämä vaatimus vieruskollokaatteja laajemmasta havaintoalueesta

ei osoittaudu ehkä kovinkaan aiheelliseksi. Vanhatalo perustelee mahdollisimman laajaa havaintoaluetta suomen kielen vapaalla sanajärjestyksellä, mikä on tulkittavissa siten, että kollokationaalisen valinnan muodostavat kollokaatit löytyisivät (suhteellisen) vapaasti noodin ympäristöstä. Omassa työssäni käyttämien astemääritteiden valossa näyttää siltä, että tyypilliset kollokaatit erottuvat kuitenkin tietystä positiosta. Näyttää siis siltä, että varsin pieneenkin alueeseen keskittyen voidaan tehdä pitäviä havaintoja synonyymisten sanojen ympäristön järjestelmällisistä leksikaalisista eroista. Tulevassa tutkimuksessa on kuitenkin selvitettävä, koskevatko tähän mennessä esiin tulleet havainnot synonyymeja yleisemminkin. Muiden kuin kollokationaalisten valintojen kuvaamisessa havaintoalueen laajuus on puolestaan taas olennainen piirre.

Yhdeksi uudeksi kolligationaaliseksi piirteeksi erottui analyysissa lausepituus. Lausepituuden analyysi tuli työssä ajankohtaiseksi kolligaattianalyysin jälkeen, kun selvisi, että lauseasemakolligaatti osoittautui usein yhdeksi synonyymit toisistaan erottavaksi kolligaatiksi sanaluokkakolligaattien rinnalla. Lausepituuden testaaminen osoittikin synonyymien eroavan tämän piirteen osalta toisistaan, ja lausepituuskolligaatti oli myös yksi niistä muuttujista, joilla pystyttiin kuvaamaan alkuperäis-suomen ja käännösuomen eroa. Kolligaattianalyysia olisi mahdollista laajentaa edelleen. Erityisesti kaunokirjallisia tekstejä analysoitaessa voisi yhdeksi mahdolliseksi kolligaattimuuttujaksi esittää suoran kerronnan (vuorosanojen) ja muun kerronnan vaikutusta synonyymien käyttöön. Silmämääräisesti aineistoa analysoitaessa näyttää siltä, että tarkastelluista määritteistä *oikein* tuntuisi viihtyvän usein juuri vuorosanoissa, kuten toivotuksissa (– *Kiitos nyt sitten vain oikein paljon*). Toinen tutkimisen arvoinen olisi mahdollisesti myös lausetyyppimuuttujan merkitys. Tähän viittaavia tuloksia nousikin jo esiin esimerkiksi luvussa 6.2.2, jossa *hyvin*-määritteen käyttö tila- ja luonnehtivissa lauseissa oli omiaan lisäämään OLLA-kollokaatin määrää sen kontekstissa ja erottamaan määritteen synonyymeistaan.

Kääntämisen aiheuttamia kontekstuaalisen valinnan muutoksia tarkasteltiin tutkimuksessa aikaisempaa laajemmin. Aiempi tutkimushan on kohdistanut huomionsa nimenomaan käännöskielen leksikaalisten myötäesiintymien erilaisuuteen (Mauranen 2000b). Tutkimuksessa selvisi, että käännösuomen leksikon kontekstuaalisia epätyypillisiä rakenteita tavataan erityisesti juuri leksikaalisissa myötäesiintymissä ja että kontekstuaalisen merkitysyksikön semanttiset ominaisuudet sallivat vähän muutoksia käännettäessä. Kenny (2001: 170–173) on tosin löytänyt saksan ja englannin välillä tehdyistä käännöksistä tapauksia, joissa leksikaalisen yksikön semanttiset preferenssit ja prosodiat ovat muuttuneet käännettäessä. Hänen tutkimuksessaan tarkastelun kohteena olivat kuitenkin yksittäiset käännöstapaukset kahden eri kielen välillä päinvastoin kuin käsillä olevassa tutkimuksessa, jossa etsittiin yleisiä, useisiin esiintymiin perustuvia suuntauksia yhden kielen eri muodoissa, joskin vain kolmen esimerkkileksemin avulla. Se, että esimerkiksi astemääritteiden semanttiset piirteet eivät kokonaisuutena osoita käännösuomen kontekstuaalisia preferenssejä erilaisiksi, ei tietystikään poista sitä mahdollisuutta, että käännösaineistossa esiintyisi sellaisia yksittäisiä rakenteita, jotka poikkeavat yleisistä trendeistä.

Tutkimus antaa myös vihjeitä siitä, että kontekstin kieliopilliset (kolligaatiot) ja semanttiset piirteet ovat jähmeitä muuttumaan käännösprosessissa. Tämä olisi si-

nänsä järkeenkäypää, koska prosodisia ominaisuuksia pidetään yhtenä lekseemien valintaa ohjaavana tekijänä. Tässä työssä pääasiallisesti tarkasteltu sanaluokkakolligaattien valinta ei myöskään voine hevin olla käännöksissä erilainen, koska näiden kolligaattien määrä on hyvin rajallinen. Toisaalta tuli ilmi, että lausepituus näyttää vaihtelevan kielimuotojen kesken. Selitysmahdollisuuksia voisi löytyä ainakin kaksi: ensinnäkin Laviosan (1996) oletus käännösten tyylillisestä yksinkertaistumisesta, joka todellistuisi lyhyempinä lauseina, voisi selittää englannista suomennettujen tekstien lyhyet kontekstilauseet. Toisaalta tätä ei kuitenkaan tue toisesta käännösaineistosta saatu tulos, jonka mukaan kontekstilauseet ovat käännöksissä yhtä pitkiä kuin alkuperäissuomessakin. Tällöin selitys voisi löytyä englannin kielen vastaavien astemääritteitä sisältävien lauseiden pituudesta, mutta tämä oletus jää tässä tutkimuksessa spekulatiiviseksi vertailumateriaalin puuttuessa.

Epätypilliset kontekstuaaliset rakenteet lienevät juuri niitä, joihin käännöstekstin vastaanottaja kiinnittää lukiessaan huomionsa. Kuinka poikkeavia rakenteiden tulisi olla, jotta ne hyppäisivät tekstistä esiin, on tutkimatonta aluetta. Mahdollisesti juuri semanttisten piirteiden rikkomisen saisi tekstin tuntumaan jollakin määrittelemättömällä tavalla oudolta. Huomiota herättävää voinee olla myös kiinteitä, ennustettavia leksikaalisia myötäesiintymiä (jopa rajoituksia) rikkova käyttö, olkoon tästä esimerkkinä aiemmin työssä (luku 5.3.1) mainittu *tuiki* <TAVALLINEN>, jota voisi verrata vaikkapa kombinaatioon *?tuiki* <MAUTON>. Uskoisin jälkimmäisen herättävän lukijan mielenkiinnon ainakin jossain määrin epätavallisuutensa vuoksi.

Mutta entäpä tässä tutkimuksessa esiin tulleet leksikaaliset myötäesiintymät? Huomaisiko lukija epätypilliseksi kollokaatiot *hyvin* <ERIKOINEN, *harvoin*>, *kovin* <*vähän, kaukana*> ja *oikein* <HAUSKA, *paljon*>. Herättäisivätkö ne miettimään käännösuomen epätavallisia sanavalintoja? Luultavasti eivät. Työssä esiin tulleet tapaukset lienevätkin sellaisia, joita ei pystytä kovinkaan helposti havaitsemaan ilman laajaa aineistoa ja tilastollisia menetelmiä. Tässä suhteessa ilmiöstä voisikin käyttää Karhun (1997) ilmausta *piilopoikkeama*, koska esiintymät ovat lukijalta, ja tutkijaltakin, piilossa ilman oikeanlaisia havaintomenetelmiä. Käsite on kuitenkin varattu jo aiemmin käytettäväksi yhteyksissä, joissa esitellään kartettaviksi luokiteltuja käännöskielen ominaisuuksia. Tutkimuksessa esiin tulleet erot onkin mielestäni nähtävä yhtenä sellaisena piirteenä, joka tuo käännöksiin jonkinlaisen käännöskielen tunnun ja jonka syytä vastaanottaja ei pysty yksiselitteisesti selittämään.

Tutkimuksessa käytetyn aineiston jakaminen kolmeen osakorpukseen antoi mahdollisuuden tarkastella, poikkeako yhdestä lähdekielestä käännetty kieli siitä yleiskuvasta, joka tulee ilmi, kun tarkastellaan yleensä käännöskieltä. Lähdekielen vaikutusta tukevia havaintoja saatiin esiin runsaanlaisesti. Se, missä määrin yhden lähdekielen vaikutuksen puolesta puhuvat tulokset johtuvat nimenomaan englannista, joka oli yksilähdekielisen korpuksen lähdekielenä, ei selviä tässä tutkimuksessa. Tämä johtuu yhtäältä siitä, että käytettävissä ei ole vastaavan kaltaista ja samoilla menetelmillä hankittua analysoitua tietoa englanninkielisistä rakenteista, ja toisaalta siitä, että aineisto ei sisällä käännösten lähdetekstejä. Toisaalta vaikka tutkittujen kolmen astemääritteen lähdekielisiä vastineina pyrittäisiinkin tutkimaan englannista, tulisi eteen kääntämisen eri mahdollisuuksista kumpuava moninaisuuden ongelma: tässä työssä tarkastelluilla synonyymeilla voi olla hyvin monenlaisia vastineita lähdeteksteissä – ja mahdollista on myös se, että määritteellä ei edes ole

suoraa vastinetta lähdetekstissä – jolloin on lähes mahdotonta tietää, minkä lähdekielisten astemääritteiden stimulusvaikutusta olisi tarkasteltava. Lisäksi huomautettakoon, että englannin valinta tämän työn yksilähdekielisen korpuksen lähdekieleksi ei johtunut pelkästään englannista sinänsä vaan tutkimuskorpuksen ulkopuolisista tekijöistä. Tältä osin tutkimus herättää mielenkiintoisen kysymyksen: onko samankaltaisia yhdestä lähdekielestä käännettyissä teksteissä havaittavia systemaattisia piirteitä löydettävissä aina, olipa lähdekieli mikä tahansa? Onko lähdekielen vaikutus niin yleismaailmallinen ilmiö, että yhdestä lähdekielestä käännettyt tekstit osoittavat aina epätyypillisiä rakenteita niin suhteessa alkuperäiskieleen kuin myös suhteessa käännskieleen kokonaisuutena?

Viimeiseksi haluan pohtia yksittäisten havaintoyksiköiden merkitystä tutkimuksessa. Jo varsin nopeastikin vilkaisemalla luvussa 7 esitettyjä taulukoita astemääritteiden alkuperäissuomesta poikkeavista kontekstuaalisista preferensseistä käännskieleessä (taulukot 7.10 s. 175, 7.17 s. 189, 7.24 s. 203) voi havaita, että synonyymisten määritteiden herkkyyys kontekstuaaliselle variaatiolle on erilainen. Siinä missä *hyvin*-määritteen kontekstissa piirteet näyttävät yli- tai aliedustuvan kaikilla tarkastelluilla tasoilla, *kovin*- ja *oikein*-määritteiden epätyypilliset piirteet ovat suomenoksissa huomattavasti vähäisemmät. Tämä suuntaus näkyy kaikissa kontekstuaalisissa piirteissä. Näyttää siltä, että *hyvin*-määrite on jonkinlainen yleisastemäärite, joka voi esiintyä hyvin monenlaisissa kontekstuaalisissa rakenteissa ja jonka lähiympäristössä on myös variaation (etenkin kvantitatiivisen variaation) mahdollisuus suhteellisen suuri. *kovin* ja *oikein* näyttävät sen sijaan olevan synonyymiaan kontekstispesifisempiä määritteitä, jotka eivät myöskään salli suurta vaihtelua ympäristössä käännskieleessä.

Havainto nostaa esiin kysymyksen siitä, mikä merkitys havaintoyksiköillä on kontekstuaalisia piirteitä tutkivassa ja käännskieleltä kuvaamaan pyrkivässä tutkimuksessa. Jos näinkin abstraktit synonyymiset ilmaukset kuin astemääritteet voivat antaa käännskielen kuvauksessa osittain toisistaan poikkeavan tuloksen, on se merkki siitä, että itse havaintoyksiköiden valinta voi vaikuttaa esiin tuleviin tuloksiin. Tässä mielessä tutkimukseni antia käännsuomen kannalta on yhtä hyvin se, millaiset leksikaaliset ominaisuudet voivat olla käännskieleessä erilaisia kuin alkuperäiskieleessä, kuin sekin, millaisen kuvan synonyymisten ilmausten keskinäinen suhde antaa käänntystä kielimuodosta.

Tutkimuksen ote oli vahvasti kontekstualisoiva. Synonymian luonnetta pyrittiin selvittämään lähimerkityksisten havaintoyksiköiden kontekstuaalisten ja erityisesti kontekstuaalisten piirteiden avulla. Tällöin lekseemien merkitystä lähestyttiin niiden käyttöympäristön kautta. Myös käännsuomen kuvauksessa meneteltiin samoin: käännsuomen leksikaalisia epätyypillisyyksiä tavoiteltiin nimenomaan käyttöympäristöä tutkimalla. Käännskielen analyysin kannalta keskittyminen yhden kielimuodon piirteisiin ei kuitenkaan riitä, vaan kielimuodosta tavattavia piirteitä on pystyttävä vertailemaan toiseen kielimuotoon, jotta tyypilliset ja epätyypilliset ominaisuudet ovat havaittavissa. Tämän tutkimusotteen mahdollisti se, että käännsuomi asetettiin alkuperäissuomen rinnalle. Näin ollen tutkimuskohteena oli kaksi suomen kielen muotoa, joilla on erilainen syntyprosessi. Mikäli tutkimuksessa esiin tullutta lähdekielen vaikutusta käännskielen luonteeseen halutaan monijärjestelmäteorian yhteydessä korostaa, voitaisiin englannista käännettyä kieltä pitää vuo-

rostaan yhtenä käännöskielen muotona, koska sillä näyttää olevan samankaltaisia erottavia ominaisuuksia suhteessa kokonaiskuvaan käännöskielestä kuin oli havaittavissa käännöskielen ja alkuperäissuomenkin välillä. Käännöskielen sisälle voitaisiin siten ajatella useita lähdekielispesifisiä muotoja, josta käännöskieli muodostuu – sen lisäksi että käännöskielellä saattaa itsessään olla kaikki sen variantit läpäiseviä ominaisuuksia, joita on tutkittu käännösuniversaaleina. Se missä määrin näin hienojakoinen erittely on järkeenkäypää ja mahdollista, ei saa tässä tutkimuksessa vastausta.

Tutkimukseni asettuu aiemman käännössuomea käsitelleen tutkimuksen rinnalle valottaen käännöskielen leksikaalisten ominaispiirteiden luonnetta. Kun analyysiani ajatellaan nimenomaan lekseemilähtöisesti, tutkimukseni ei pysty kertomaan kovin monen ilmauksen käytöstä käännöksissä. Samoin analyysin rajoittaminen yhteen tekstilajiin asettaa tulosten yleistettävyydelle rajoituksensa. Tutkimuksen lähtökohdista olikin laajentaa käännöskielestä aiemmin tehtyä kotekstianalyysia kollokaatteja abstraktimmalle tasolle, jolloin saadaan entistä tarkempi kuva niistä kotekstuaalisista valinnoista, jotka voivat olla käännöskielessä alkuperäiskielestä poikkeavia. Keskittyminen frekvenssianalyysihin ja kollokaatiosuhteisiin olisi mahdollistanut hyvinkin laajan sanakirjon tarkastelun, mutta silloin olisi jäänyt tavoittamatta tässä tutkimuksessa saatu tieto erilaisten syntagmaattisten suhteiden muutosherkkyydestä. Tällaisten käännöskielen piirteiden löytäminen – kuten ylipäätään käännöskielen erilaisten piirteiden havaitseminen – on tärkeää etenkin sellaisessa kulttuurissa, jossa käännöskieltä käytetään runsaasti. Koska suomennosten suomi on vankka osa suomalaista kielijärjestelmää, ovat sen piirteet itsearvoisesti jo sellaisia tutkimuskohteita, joita ei voi suomen kielen tutkimuksessa jättää käsittelemättä. Suomentamisen tuottamien uusien, kieltä ”venyttävien” piirteiden analysointia olisi myös harjoitettava mahdollisen pysyvemmän muutosvaikutuksen tähden: mikä onkaan ollut suomentamisen vaikutus esimerkiksi kirjakielen kehitykseen ja muutokseen? Millä tavalla piirteet, joiden on oletettu olevan kääntämisen lainalaisuuksia (universaalihypoteesit) – kuten käännöskielen yksinkertaisuus, eksplisiittisyys ja konventionaalisuus – ovat mahdollisesti olleet läsnä ja vaikuttaneet suomen kirjakielen kehitykseen? Tähän liittyvän keskustelun avauksia onkin jo nähtävissä Paloposken (2002, tulossa) ja Murasen (2004b) puheenvuoroissa. Käännösprosessin huomioon ottava diakroninen tutkimus olisi omiaan täydentämään kielen kuvausta ja tällä hetkellä jo runsaastikin harjoitettavaa synkronista tutkimusta – muistettakoon myös, että diakronisen muutoksen kuvaus on yksi monijärjestelmäteorian peruslähtökohdista. Tältä kannalta katsottuna suomen kielen kuvaus voisi saada uusia näkökulmia.

LÄHTEET

Aineistolähteet

Verrannollinen kaunokirjallisuuden korpus

AS-aineisto:

CARLSON, KRISTIINA 1999: Maan ääreen. Otava.
 HAAKANA, ANNA-LIISA 1997: Se mitä et muista. Gummerus.
 HASSINEN, PIRJO 1998: Viimeinen syli. Otava.
 HOLAPPA, PENTTI 1998: Ystävän muotokuva. WSOY.
 HOTAKAINEN, KARI 1995: Syntisäkki. WSOY.
 HUOVI, HANNELE 1994: Paratiisipuku ja muita naisten unia. Tammi.
 HUOVINEN, VEIKKO 1999: Porsaan paperit. WSOY.
 KAURANEN, ANJA 1996: Syysprinssi. WSOY.
 KROHN, LEENA 1995: Ettei etäisyys ikävöisi. WSOY.
 KYLÄTASKU, JUSSI 2000: Jalmarin toinen tuleminen. WSOY.
 LIUKKONEN, TERO 1996: Häätanhu. Gummerus.
 MALKAMÄKI, SARI 1997: Sunnuntaina kahdelta. Otava.
 MÄKELÄ, JUHANI 1995: Mämmi Ltd. WSOY.
 MÄKI, REIJO 1996: Tatuoitu taivas. Otava.
 MÖRÖ, MARI 1998: Kiltin yön lahjat. WSOY.
 ORANEN, RAIJA 1997: Puhtaat valkeat lakanat. Gummerus.
 RANIVAARA, JORMA 1999: Rouva Korhosen tapaus. WSOY.
 RINTALA, PAAVO 1996: Faustus. Otava.
 SALMINEN, ARTO 1998: Varasto. WSOY.
 SIEKKINEN, KEIJO 1997: Papin poika ja pappi. Gummerus.
 SIEPPI, PAULA 1996: Aurinko maassa. Gummerus.
 SYRJÄ, JUHANI 1996: Minotauros. Epätieteellinen kirjoitus. Gummerus.
 TAMMINEN, PETRI 1997: Miehen ikävä. Otava.
 TERVO, JARI 1995: Pyhiesi yhteyteen. WSOY.
 TUOMINEN, TAIJA 2000: Tiikerihai. WSOY.
 TUURI, ANTTI 1996: Suomi elää metsistään. Otava.

KSM-aineisto:

ATXAGA, BERNAND 1999: Tuolla taivaalla (Zeru horiek, espanjasta suom. Tarja Roinila). Tammi.
 CRICHTON, MICHAEL 1997: Suora lento (Airframe, englannista suom. Jaakko Kankaanpää). Otava.
 FREDRIKSSON, MARIANNE 2000: Marian evankeliumi (Enligt Maria Magdalena, ruotsista suom. Laura Jänisniemi). Otava.

- GORDIMER, NADINE 1996: Ei seuraa matkalleni (None to accompany me, englannista suom. Kersti Juva). WSOY.
- GRASS, GÜNTER 1996: Avarammille aloille (Ein weites Feld, saksasta suom. Oili Suominen). Tammi.
- ISEGAWA, MOSES 1999: Abessinialaiset kronikat (Abessijnse kronieken, hollannista suom. Kirsti Nikkanen). WSOY.
- KROSS, JAAN 1997: Mesmerin piiri (Mesmeri ring, virosta suom. Juhani Salokannel). WSOY.
- 1999: Paikallaanlento (Paigallend, virosta suom. Jouko Vanhanen). WSOY.
- MAKINE, ANDREI 1997: Ranskalainen testamentti (Le testament francais, ranskasta suom. Annikki Suni). WSOY.
- MÉSZÖLY, MIKLÓS 1999: Urheilijan kuolema (Az atleta halala, unkarista suom. Anna Tarvainen). Taifuuni.
- ROSENHOLM, ARJA – SALMINEN, EILA (toim.) 1997: Puhu, Maria! Uutta venäläistä naiskirjallisuutta (venäjistä suom. P. Autio, M.-L. Jaakkola, A. Pikkupeura, A. Rosenholm, E. Salminen). Tammi.
- SHMELJOV, NIKOLAI 1998: Matkalla minä sairastuin (V puti – zanemog, venäjistä suom. Tuula Jormalainen). Puijo.
- ULLMAN, LINN 1999: Ennen unta (För du sovner, norjasta suom. Tarja Teva). WSOY.

KSE-aineisto:

- ATWOOD, MARGARET 1997: Nimeltään Grace (Alias Grase, suom. Kristiina Drews). Otava.
- BELLE, JENNIFER 1997: Se on menoa (Going Down, suom. Kaijamari Sivill). Otava.
- BRODSKY, JOSEPH 1997: Keräilijän kappale (On Grief and Reason, suom. Kalevi Nytytäjä). Tammi.
- CARD, ORSON S. 1998: Punainen profeetta. Alvinin toinen kirja (The Red Prophet, suom. Jaakko Kankaanpää). Otava.
- COETZEE, J. M. 1999: Poikavuodet. Kohtauksia syrjäisestä elämästä (Boyhood. Scenes from provincial Life, suom. Seppo Loponen). Otava.
- DENES, MAGDA 1998: Palavat linnat. Sota lapsen silmin (Castles burning. A Child's Life in War, suom. Heikki Eskelinen). Otava.
- PHILLIPS, JAYNE 1995: Suojelus (Shelter, suom. Kersti Juva). Tammi.
- STEEL, DANIELLE 1997: Viisi päivää Pariisissa (Five Days in Paris, suom. Kaisa Luntinen). WSOY.
- STEVENSON, ROBERT 1997: Kävelyretkistä (Roads, suom. Alice Martin). Like.
- TAN, AMY 1996: Sata salattua aistia (The Hundred Secret Senses, suom. Eva Siikarla). WSOY.
- WILSON, ROBERT 1998: Eureka Street, Belfast (Eureka Street, suom. Markku Päkikilä). Otava.

Kirjallisuus

- AHLMAN, ERIK 1933: Adverbeista. – *Virittäjä* 37 s. 137–159.
- ALLERTON D. J. 1984: Three (or four) levels of word cooccurrence restriction. – *Lingua* 63 s. 17–40.
- ALTENBERG, BENGT 1991: Amplifier collocations in spoken English. – Stig Johansson – Anna-Brita Stenström (toim.), *English Computer Corpora. Selected Papers and Research Guide* s. 127–147. Mouton de Gruyter, Berlin.
- ARPE, ANTTI 2001: Corpus-Based Observations on the Use and Occurrence of Inflected Word Forms in Finnish – the Case of Synonyms. – Jussi Niemi – Janne Heikkinen (toim.), *Nordic and Baltic Morphology: Papers from a NorFA Course*. Tartu, June 2000. *Studies in Languages*, University of Joensuu 36 s. 1–14.
- 2002: The usage patterns and selectional preference of synonyms in a morphologically rich language. – Annie Morin – Pascal Sébillot (toim.), *Proceedings of the 6th International Conference on Textual Data Statistical Analysis (JADT 2002)*. Vol. 1 s. 21–32. Inria, Rennes.
- ASTON, GUY – BURNARD, LOU 1998: *The BNC Handbook. Exploring the British National Corpus with SARA*. Edinburgh University Press.
- BAKER, MONA 1992: In *Other Words: A Coursebook on Translation*. Routledge, London.
- 1993: *Corpus Linguistics and Translation Studies. Implications and Applications*. – Mona Baker – Gill Francis – Elena Tognini-Bonelli (toim.), *Text and Technology. In Honour of John Sinclair* s. 233–250. Benjamins, Amsterdam.
- 1995: *Corpora in Translation Studies. An Overview and Some Suggestions for Future Research*. – *Target* 7(2) s. 223–243.
- 1996: *Corpus-based translation studies. The challenges that lie ahead*. – Harold Somers (toim.), *Terminology, LSP and Translation. Studies in Language Engineering in Honour of Juan C. Sager* s. 175–186. Benjamins, Amsterdam.
- 1999: *The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators*. – *International Journal of Corpus Linguistics* 4(2) s. 281–298.
- BARNBROOK, GEOFF 1996: *Language and Computers. Practical Introduction to the Computer Analysis of Language*. Edinburgh University Press.
- BASSNETT, SUSAN 1998: *The Translation Turn in Cultural Studies*. – Susan Bassnett – André Lefevere (toim.), *Constructing Cultures. Essays on Literary Translation* s. 123–140. *Multilingual Matters*, Clevedon.
- BEAUGRANDE, ROBERT de 1999: *Reconnecting Real Language with Real Texts. Text Linguistics and Corpus Linguistics*. – *International Journal of Corpus Linguistics* 4(2) s. 243–259.
- BERNARDINI, SILVIA – ZANETTIN, FEDERICO 2004: *When is a universal not a universal? Some limits of current corpus-based methodologies for the investigation of translations universals*. – Anna Mauranen – Pekka Kujamäki (toim.), *Translation Universals. Do They Exist?* s. 51–62.

- BERRY-ROGGHE, GODELIEVE 1973: *The Computation of Collocations and Their Relevance in Lexical Studies*. – A. J. Aitken – R. W. Bailey – N. Hamilton-Smith (toim.), *The Computer and Literary Studies* s. 103–112. Edinburgh University Press.
- BIBER, DOUGLAS – CONRAD, SUSAN 1999: *Lexical Bundles in Conversation and Academic Prose*. – Hilde Hasselgård – Signe Oksefjell (toim.), *Out of Corpora. Studies in Honour of Stig Johansson* s. 181–190. Rodopi, Amsterdam.
- – CONRAD, SUSAN – REPPEN, RANDI 1998: *Corpus Linguistics. Investigation Language Structure and Use*. Cambridge University Press.
- BLUM-KULKA, SHOSHANA – LEVENSTON EDDIE A. 1983: *Universals of Lexical Simplification*. – Claus Faerch – Gabriele Kasper (toim.), *Strategies in Interlanguage Communication* s. 119–139. Longman, London.
- BOLINGER, DWIGHT 1972: *Degree Words*. Mouton, The Hague.
- BOWKER, LYNNE 2000: *Towards a Methodology for Exploiting Specialized Target Language Corpora as Translation Resources*. – *International Journal of Corpus Linguistics* 5(1) s. 17–52.
- BUTLER, CHRISTOPHER 1985: *Statistics in Linguistics*. Blackwell, Oxford.
- BÄCKLUND, ULF 1973: *The collocation of Adverbs of Degree in English*. *Acta Universitatis Upsaliensis. Studia Anglistica Upsaliensia* 13.
- CARTER, RONALD 1987: *Vocabulary. Applied Linguistic Perspectives*. Allen & Unwin, London.
- CD-PS 1997 = CD-Perussanakirja 1997. *Kotimaisten kielten tutkimuskeskuksen julkaisuja* 94. *Kotimaisten kielten tutkimuskeskus – Edita – Lingsoft*, Helsinki.
- CHAFE, WALLACE 1992: *The Importance of Corpus Linguistics to Understanding the Nature of Language*. – Jan Svartvik (toim.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991* s. 79–97. Mouton de Gruyter, Berlin.
- CHESTERMAN, ANDREW 1998: *Finnish Tradition*. – Mona Baker (toim.), *Routledge Encyclopedia of Translation Studies* s. 401–408. Routledge, London.
- 2000: *Kriitikko ja käännösefektit*. – Outi Paloposki – Henna Makkonen-Craig (toim.), *Käännöskirjallisuus ja sen kritiikki* s. 62–78. AKO, Helsinki.
- 2001: *Beyond the particular. Esitelmä konferenssissa Translation Universals – Do They Exist?*, Savonlinna, 19.–20.10.2001.
- 2004: *Beyond the particular*. – Anna Mauranen – Pekka Kujamäki (toim.), *Translation Universals. Do They Exist?* s. 33–49.
- CHURCH, KENNETH – GALE, WILLIAM – HANKS, PATRICK – HINDLE, DONALD 1991: *Using Statistics in Lexical Analysis*. – Uri Zernik (toim.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon* s. 115–164. Lawrence Erlbaum Associates, Hillsdale.
- – HANKS, PATRICK 1990: *Word Association Norms, Mutual Information, and Lexicography*. – *Computational Linguistics* 16(1) s. 22–29.

- – MERCER, ROBERT 1993: Introduction to the Special Issue on Computational Linguistics Using Large Corpora. – *Computational Linguistics* 19(1) s. 1–24.
- CLEAR, JEREMY 1993: From Firth Principles. Computational Tools for the Study on Collocation. – Mona Baker – Gill Francis – Elena Tognini-Bonelli (toim.), *Text and Technology*. In Honour of John Sinclair s. 271–292. Benjamins, Amsterdam.
- CRUSE, D. ALAN 1986: *Lexical Semantics*. Cambridge University Press.
- 2000: *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford University Press.
- CRYSTAL, DAVID 1985: *A Dictionary of Linguistics and Phonetics*. Basil Blackwell, Oxford.
- DIXON, R. M. W. 1977: Where have all the adjectives gone? – *Studies in Language* 1 (1) s. 19–80.
- ESKOLA, SARI 2002: Syntetisoivat rakenteet käännösuomessa. Suomennetun kaunokirjallisuuden ominaispiirteiden tarkastelua korpusmenetelmillä. Joensuu yliopiston humanistisia julkaisuja 30.
- EVEN-ZOHAR, ITAMAR 1978: The Position of Translated Literature within the Literary Polysystem. – James S. Holmes – José Lambert – Raymond van den Broeck (toim.), *Literature and Translation. New Perspectives in Literary Studies* s. 117–127. Acco, Leuven.
- 1990 [1979]: Polysystem Theory. – *Polysystem Studies. Poetics Today* 11 (1) s. 9–26.
- 1990: The "Literary System". – *Polysystem Studies, Poetics Today* 11 (1) s. 27–44.
- FAWCETT, PETER 1998: Linguistic approaches. – Mona Baker (toim.), *Routledge Encyclopedia of Translation Studies* s. 120–125. Routledge, London.
- FIRTH, J. R. 1957: *Papers in Linguistics 1934–51*. Oxford University Press.
- 1968 [1957]: A Synopsis of Linguistic Theory, 1930–55. – F. R. Palmer (toim.), *Selected Papers of J. R. Firth 1952–59* s. 168–205. Longman, London.
- FRAWLEY, WILLIAM 1984: Prolegomenon to a Theory of Translation. – William Frawley (toim.), *Translation: Literary, Linguistic and Philosophical Perspectives* s. 159–175. Associated University Presses, London.
- 1992: *Linguistic Semantics*. Lawrence Erlbaum Associates, Hillsdale.
- GELLERSTAM, MARTIN 1986: Translationese in Swedish Novels Translated from English. – Lars Wollin – Hans Lindquist (toim.), *Translation Studies in Scandinavia. Proceedings from The Scandinavian Symposium on Translation Theory (SSOTT) II, Lund 14–15 June, 1985* s. 88–95. Liber Förlag, Malmö.
- GRANGER, SYLVIANE – WYNNE, MARTIN 2000: Optimising Measures of Lexical Variation in EFL Learner Corpora. – John M. Kirk (toim.), *Corpora Galore. Analyses and Techniques in Describing English. Papers from the Nineteenth International Conference on English Language Research on Computerised Corpora (ICAME 1998)* s. 249–257. Rodopi, Amsterdam.

- GREENBAUM, SIDNEY 1970: *Verb-intensifier Collocations in English. An Experimental Approach*. Mouton, The Hague.
- HAKANEN, AIMO 1973: *Adjektiivien vastakohtasuhteet suomen kielessä*. SKS, Helsinki.
- HAKULINEN, AULI 1975: *Sanajärjestyksen eri tehtävistä*. – *Virittäjä* 79 s. 85–94.
- – KARLSSON, FRED 1979: *Nykysuomen lauseoppia*. SKS, Helsinki.
- – KARLSSON, FRED – VILKUNA, MARIA 1996 [1980]: *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus*. Helsingin yleisen kielitieteen laitoksen julkaisuja 6. Helsingin yliopistopaino.
- – OJANEN, JUSSI 1976: *Kielitieteen ja fonetiikan termistöä*. SKS, Helsinki.
- HAKULINEN, LAURI 1969: *Suomen sanaston käänntslainoja*. SKS, Helsinki.
- HALLIDAY, M. A. K. 1966: *Lexis as a Linguistic Level*. – C. Bazell – J. Catford – M. A. K. Halliday – R. Robins (toim.), *In Memory of J. R. Firth* s. 148–162. Longman, London.
- – HASAN, RUQAIYA 1976: *Cohesion in English*. Longman, London.
- HALVERSON, SANDRA 1998: *Translation Studies and Representative Corpora: Establishing Links between Translation Corpora, Theoretical/Descriptive Categories and a Conception of the Object of Study*. – *Meta* 43(4) s. 494–514.
- HASKEL, PEGGY 1971: *Collocations as a measure of stylistic variety*. – Roy A. Wisbey (toim.), *The Computer in Literary and Linguistic Research. Papers from a Cambridge Symposium* s. 159–168. Cambridge University Press.
- HEIKKINEN, VESA 1999: *Ideologinen merkitys kriittisen tekstintutkimuksen teoriassa ja käytännössä*. SKS, Helsinki.
- – HIIDENMAA, PIRJO – TIILILÄ, ULLA 2000: *Teksti työnä, virka kielenä*. Gaudeamus, Helsinki.
- – LEHTINEN, OUTI – LOUNELA, MIKKO 2001: *Kuvia kirjoitetusta suomesta*. – *Kielikello* 3/2001 s. 12–15.
- HELLEMANN, JARL 1970: *Suomenoskirjallisuus*. – Pekka Tarkka (toim.), *Suomen kirjallisuus VIII. Kirjallisuuden lajeja* s. 418–484. Otava, Helsinki.
- HERMANS, THEO 1999: *Translation in Systems. Descriptive and System-oriented Approaches Explained*. St. Jerome, Manchester.
- HOEY, MICHAEL 1997: *From concordance to text structure: new uses for computer corpora*. – Barbara Lewandowska-Tomaszczyk – Patrick Melia (toim.), *PALC'97. Applications in Language Corpora Proceedings* s. 2–23. Łódź University Press.
- HOGG, RICHARD M. 1977: *English quantifier systems*. NHC, Amsterdam.
- HS = *Helsingin Sanomat*. Sanoma Osakeyhtiö, Helsinki.
- HUMBLEY, JOHN – KOPY, GEOFFREY S. – WRIGHT, SUE E. 1999: *English Terminology*. – Jean Delisle, Hannelore Lee-Jahnke – Monique C. Cormier (toim.), *Terminologie de la traduction – Translation terminology – Terminología de la traducción – Terminologie der Übersetzung* s. 107–212. Benjamins, Amsterdam.
- ITKONEN-KAILA, MARJA 1997: *Mikael Agricolan Uusi testamentti ja sen erikieliset lähtötekstit*. Suomi 184. SKS, Helsinki.

- JACOBSON, SVEN 1964: Adverbial positions in English. Studentbok, Stockholm.
- JALONEN, OLLI 1999: Käännöskirjallisuuden uudistuminen. – Pertti Lassila (toim.), Suomen kirjallisuushistoria 3. Rintamakirjeistä tietoverkkoihin s. 148–155. SKS, Helsinki.
- JANTUNEN, JARMO H. 1995: Em minä sitä varsin teheny! – Suomen kielen *varsin*-lekseemin tarkastelua puhe- ja kirjakieleen kuuluvana semanttisena ja syntaktisena yksikkönä. Pro gradu -tutkielma. Joensuun yliopisto.
- 2000: What Can Corpora Tell Us about Translated Language? A Comparable Corpus of Finnish in Use for Making Hypotheses. Esitelmä konferenssissa Research Models in Translation Studies 28–30 April 2000. Manchester, UK.
- 2001a: Synonymity and Lexical Simplification in Translations: A Corpus-based Approach. – Across Languages and Cultures 2(1) s. 97–112.
- 2001b: *Tärkeä seikka ja keskeinen kysymys*: Mitä korpuslingvistinen analyysi paljastaa lähisynonyymeista? – Virittäjä 105 s. 170–192.
- 2002: Comparable Corpora in Translation Studies: Strengths and Limitations. – SKY Journal of Linguistics 15 s. 105–117.
- 2003: Suomennosten kieliasu – oletuksista kohti realiteetteja. Esitelmä KäTu2003-symposiumissa 4.–5.4.2003, Turku.
- 2004: Let's count mistakes! - A corpus-based view of linguistic correctness in translations. – Marja Nenonen (toim.), Papers from the 30th Finnish Conference of Linguistics, Joensuu, May 15-16, 2003. Studies in Languages, University of Joensuu 39 s. 55–60.
- – ESKOLA, SARI 2002: Käännössuomi kielivarianttina: syntaktisia ja leksikaalisia erityispiirteitä. – Virittäjä 106 s. 184–207.
- JOHANSSON, STIG 1980: The LOB Corpus of British English Texts: Presentation and Comments. – ALLC Journal 1(1) s. 25–36.
- 1993: "Sweetly oblivious": some aspects of adverb-adjective combinations in present-day English. – Michael Hoey (toim.), Data, Description, Discourse. Papers on the English Language in honours of John Sinclair s. 39–49. HarperCollins, London.
- 1998: On the role of Corpora in Cross-linguistic Research. – Stig Johansson – Signe Oksefjell (toim.), Corpora and Cross-linguistic Research. Theory, Method, and Case Studies s. 3–24. Rodopi, Amsterdam.
- – LEECH, GEOFFREY – GOODLUCK, HELEN 1978: Manual of Information to Accompany the Lancaster-Olds/Bergen Corpus of British English, for Use with Digital Computers. University of Oslo.
- JONES, SUSAN – SINCLAIR, JOHN 1974: English Lexical Collocations. A Study in Computational Linguistics. – Cahiers de Lexicologie 24 s. 15–61. Didier-Larousse, Paris.
- KANGASNIEMI, HEIKKI 1997: Sana, merkitys, maailma. Katsaus leksikaalisen semantiikan perusteisiin. Finn Lectura, Helsinki.
- KARHU, ANNA 1997: Suomennosten nominaalilausekkeiden piilopoikkeamat. – Anna Mauranen – Tiina Puurtinen (toim.), Translation – Acquisition – Use. AFinLAN vuosikirja 1997 s. 71–87. Suomen soveltavan kielitieteen yhdistys AFinLA, Jyväskylä.

- KARLSSON, FRED 1982: Suomen kielen äänne- ja muotorakenne. WSOY, Helsinki.
- 1999: Yleinen kielitiede. Helsingin yliopistopaino.
- KARVONEN, PIRJO 1995: Oppikirjateksti toimintana. SKS, Helsinki.
- KEMPPANEN, HANNU 2002: Ystävyyden tuontia venäjältä suomeen: avainsanat ideologian toteutumina käännettyissä ja alkuperäisissä historiateksteissä. – Anna Mauranen – Liisa Tiittula (toim.), Kieli yhteiskunnassa – yhteiskunta kielessä. AFinLAN vuosikirja 2002 s. 185–205. Suomen soveltavan kielitieteen yhdistys AFinLA, Jyväskylä.
- (tulossa) Avainsanoja, ideologiaa ja lojaaliutta: venäjältä suomennettujen historiatekstien korpusanalyysi. – Anna Mauranen – Jarmo H. Jantunen (toim.), Käännösuomeksi: tutkimuksia suomennosten kielestä. Tampere Studies in Language, Translation and Culture n:o 1.
- KENNEDY, GRAEME 1991: *Between and through: the company they keep and the functions they serve.* – Karin Aijmer – Bengt Altenberg (toim.), English Corpus Linguistics. Studies in Honour of Jan Svartvik s. 95–110. Longman, London.
- KENNY, DOROTHY 1998: Corpora in translation studies. – Mona Baker (toim.), Routledge Encyclopedia of Translation Studies s. 50–53. Routledge, London.
- 2000: Lexical Hide-and-Seek: Looking for Creativity in a Parallel Corpus. – Maeve Olohan (toim.), Intercultural faultlines. Research models in translation studies I. Textual and cognitive aspects s. 93–104. St. Jerome, Manchester.
- 2001: Lexis and Creativity in Translation. A Corpus-based Study. St. Jerome, Manchester.
- Kieli ja sen kieliopit 1994: Kieli ja sen kieliopit: opetuksen suuntaviivoja. Opetusministeriö, Helsinki.
- KILGARRIFF, ADAM 2001: Comparing Corpora. – International Journal of Corpus Linguistics 6(1) s. 97–133.
- KIURU, SILVA 1977: Suomen kielen kieltohakuiset verbit. SKS, Helsinki.
- KJELLMER, GÖRAN 1984: Some Thoughts on Collocational Distinctiveness. – Jan Aarts – Willem Meijs (toim.), Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research s. 163–171. Rodopi, Amsterdam.
- 1987: Aspects of English Collocations. – Willem Meijs (toim.), Corpus Linguistics and Beyond. Proceedings of the Seventh International conference on English Language Research on Computerized Corpora s. 133–140. Rodopi, Amsterdam.
- 1990: Patterns of Collocability. – Jan Aarts – Willem Meijs (toim.), Theory and Practice in Corpus Linguistics s. 163–173. Rodopi, Amsterdam.
- 1991: A mint of phrases. – Karin Aijmer – Bengt Altenberg (toim.), English Corpus Linguistics. Studies in honour of Jan Svartvik s. 111–127. Longman, London.
- KLAUDY, KINGA 1996: Concretization and Generalization of Meaning in Translation. – Marcel Thelen – Barbara Lewandowska-Tomaszczyk (toim.), Translation and Meaning Part 3. Proceedings of the Maastricht Session of

- the 2nd International Maastricht-Lodz Duo Colloquium on “Translation and Meaning” Hogeschool Maastricht s. 141–152.
- KLEIN, HENNY 1998: *Adverbs of Degree in Dutch and Related Languages*. Benjamins, Amsterdam.
- KOHN, JÁNOS 1996: What Can (Corpus) Linguistics Do for Translation? – Kinga Klaudy – José Lambert – Aniko Sohár (toim.), *Translation Studies in Hungary* s. 39–52. Scholastica, Budapest.
- KOVALA, URPO 1996: Maailmankirjallisuuden ja kansalliskirjallisuuden välissä? Suomenoskirjallisuuden asemasta 1800-luvulla. – Juhani Niemi (toim.), *Rakkaudesta lukemiseen. Suomalaisen kirjallisuusharrastuksen kehityksestä* s. 41–59. Helsingin yliopistopaino.
- 1999: Käännöskirjallisuuden vanhat ja uudet tehtävät. – Lea Rojola (toim.), *Suomen kirjallisuuden historia 2. Järkiuskosta vaistojen kapinaan* s. 299–309. SKS, Helsinki.
- KRETZCHMAR, WILLIAM – MEYER, CHARLES – INGEGNERI, DOMINIQUE 1997: Use of Inferential Statistics in Corpus Studies. – Magnus Ljung (toim.), *Corpus-based Studies in English*. ICAME 17 s. 167–177. Rodopi, Amsterdam.
- KRISHNAMURTHY, RAMESH 1996: Ethnic, racial and tribal: The language of racism? – Carmen Caldas-Coulthard – Malcolm Coulthard (toim.), *Texts and Practices. Readings in Critical Discourse Analysis* s. 129–149. Routledge, London.
- KUČERA, HENRY – FRANCIS, W. NELSON 1967: *Computational Analysis of Present-Day American English*. Brown University Press.
- KUIRI, KAIJA 2000: *Semantiikan peruskurssi*. Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
- KUIVASMÄKI, RIITTA 1996: Kuinka nuorisokirjallisuus löysi paikkansa kirjallisuusinstituutiossa. – Juhani Niemi (toim.), *Rakkaudesta lukemiseen. Suomalaisen kirjallisuusharrastuksen kehityksestä* s. 60–72. Helsingin yliopistopaino.
- KUJAMÄKI, PEKKA 2004: What happens to “unique items” in learners’ translations? ”Theories” and ”concepts” as a challenge for novices’ views on ”good translation”. – Anna Mauranen – Pekka Kujamäki (toim.), *Translation Universals. Do They Exist?* s. 187–204.
- – JÄÄSKELÄINEN, RIITTA 2001: Korpukset kääntäjän apuna. – Riitta Oitinen – Pirjo Mäkinen (toim.), *Alussa oli käännös* s. 239–252. Tampereen yliopistopaino.
- LAMBERT, JOSÉ 1997: Itamar Even-Zohar’s Polysystem Studies: An Interdisciplinary Perspective on Culture Research. – *Canadian Review of Comparative Literature* 24(1) s. 8–14.
- LAPPALAINEN, HANNA 2000: Sosiolingvistinen katsaus suomalaisnuorten nykypuhekieleen ja sen tutkimukseen. – *Virittäjä* 104 s. 74–101.
- LAVIOSA, SARA 1997: How Comparable can ‘Comparable Corpora’ be? – *Target* 9(2) s. 289–319.
- 1998: *The English Comparable Corpus. A Resource and a Methodology*. – Lynne Bowker – Michael Cronin – Dorothy Kenny – Jennifer Pearson

- (toim.), *Unity in Diversity? Current Trends in Translation Studies* s. 101–112. St. Jerome, Manchester.
- 2002: *Corpus-based Translation Studies. Theory, Findings, Applications*. Rodopi, Amsterdam.
- LAVIOSA-BRAITHWAITE, SARA 1996: *The English Comparable Corpus (ECC): A Resource and a Methodology for The Empirical Study of Translation*. Painamaton väitöskirja, UMIST, Manchester. UK.
- 1997: *Investigating Simplification in an English Comparable Corpus of Newspaper Articles*. – Kinga Klaudy – János Kohn (toim.), *Transfere Necesse Est. Proceedings of the 2nd International Conference on Current Trends in Studies of Translation and Interpreting* s. 531–540. Scholastica, Budapest.
- 1998: *Universals of Translation*. – Mona Baker (toim.), *Routledge Encyclopedia of Translation Studies* s. 288–291. Routledge, London.
- LEECH, GEOFFREY 1981: *Semantic. The Study of Meaning*. Penguin Books.
- LEINO, ANTTI – LEINO, PIRKKO 1990: *Synonyymisanasto*. Otava, Helsinki.
- LEINO, PENTTI 1990: *Sijojen asema kielenkuvauksessa*. – Pentti Leino – Marja-Liisa Helasvuo – Petri Lauerma – Urpo Nikanne – Tiina Onikki (toim.), *Suomen kielen paikallissijat konseptuaalisessa semantiikassa*. *Kieli* 5 s. 12–58. Helsingin yliopiston suomen kielen laitos.
- 1993: *Polysemia – kielen moniselitteisyys*. *Kieli* 7. Helsingin yliopiston suomen kielen laitos.
- LEPPÄJÄRVI, EILA – JÄMSÄ, TUOMO 1976: *Adjektiivit 1960-luvun Suomen lehti-, radio- ja yleispuhekielessä*. Oulun yliopiston suomen ja saamen kielen laitoksen tutkimusraportteja 7.
- LILIUS, PIRKKO 2000: *Käännöskirjallisuus 1600- ja 1700-luvun Suomessa*. – Outi Paloposki – Henna Makkonen-Craig (toim.), *Käännöskirjallisuus ja sen kritiikki* s. 5–19. AKO, Helsinki.
- LINDQUIST, HANS 1999: *Electronic Corpora as Tools for Translation*. – Gunilla Anderman – Margaret Rogers (toim.), *Word, Text, Translation. Liber Amicorum for Peter Newmark* s. 179–189. *Multilingual Matters*, Clevedon.
- LOUW, BILL 1993: *Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies*. – Mona Baker – Gill Francis – Elena Tognini-Bonelli (toim.), *Text and Technology. In Honour of John Sinclair* s. 157–176. Benjamins, Amsterdam.
- LYONS, JOHN 1966: *Firth's Theory of "Meaning"*. – C. Bazell – J. Catford – M. A. K. Halliday – R. Robins (toim.), *In Memory of J. R. Firth* s. 288–302. Longman, London.
- 1968: *Introduction to Theoretical Linguistics*. Cambridge University Press.
- 1981: *Language, Meaning and Context*. Fontana, London.
- LÄNSIMÄKI, MAIJA 1995: *Voiko sodassa saada ääsiyyskutsun? Miten kuoleminen ilmaistaan suomen kielessä*. – Maija Länsimäki (toim.), *Sana kiertää: kirjoituksia suomen kielestä* s. 63–69. Kotimaisten kielten tutkimuskeskus, Helsinki.

- MALINOWSKI, BRONISLAW 1930: The problem of meaning in primitive languages. – C. K. Odgen – I. A. Richards (toim.), *The Meaning of Meaning* s. 296–336. Routledge & Kegan Paul, London.
- MALMKJÆR, KIRSTEN 1993: Who Can Make Nice A Better Word Than Pretty? Collocation, Translation, and Psycholinguistics. – Mona Baker – Gill Francis – Elena Tognini-Bonelli (toim.), *Text and Technology*. In Honour of John Sinclair s. 213–232. Benjamins, Amsterdam.
- 1997: Punctuation in Hans Christian Andersen’s stories and in their translations into English. – Fernando Poyatos (toim.), *Nonverbal Communication and Translation. New Perspectives and Challenges in Literature, Interpretation and the Media* s. 151–162. Benjamins, Amsterdam.
- MASON, OLIVER 1997: The weight of words: an investigation of lexical gravity. – Barbara Lewandowska-Tomaszczyk – Patrick Melia (toim.), *PALC 97. Applications in Language Corpora Proceedings* s. 361–375. Łódź University Press.
- 2000: Parameters of Collocation: The Word in the Centre of Gravity. – John M. Kirk (toim.), *Corpora Galore. Analyses and Techniques in Describing English. Papers from the Nineteenth International Conference on English Language Research on Computerised Corpora (ICAME 1998)* s. 267–280. Rodopi, Amsterdam.
- MAURANEN, ANNA 1998a: Käännösuomi ja kääntämisen universaalit: tutkimus korpussaineistolla. Projektisuunnitelma, mimeo. Joensuun yliopisto.
- 1998b: Form and sense relations as seen through parallel corpora. – Wolfgang Teubert – Elena Tognini-Bonelli – Norbert Volz (toim.), *TELRI. Proceedings of the Third European Seminar “Translation Equivalence”*. Montecatini Terme, Italy, October 16–18, 1997 s. 159–173. TWC, Birmingham.
- 2000a: Strange strings in translated language. A study on corpora. Esitelmä konferenssissa *Research Models in Translation Studies 28–30 April 2000*. Manchester, UK.
- 2000b: Strange strings in translated language. A study on corpora. – Maeve Olohan (toim.), *Intercultural faultlines. Research models in translation studies I. Textual and cognitive aspects* s. 119–141. St. Jerome, Manchester.
- 2002a: Will ‘translationese’ ruin a contrastive study? – *Languages in Contrast* 2(2) s. 161–185.
- 2002b: Käännöstekstit ja genre. Esitelmä Kotikielen Seuran seminaarissa Kirjoitetun suomen lähteillä – kääntämisen teemapäivä. Helsingin yliopisto, 13.4.2002.
- 2004a: Corpora, universals and interference. – Anna Mauranen – Pekka Kujamäki (toim.), *Translation Universals. Do They Exist?* s. 65–82.
- 2004b: Historian puheenvuoro käännöskeskusteluun. – *Virittäjä* 108 s. 125–133.
- – KUJAMÄKI, PEKKA (toim.) 2004: *Translation Universals. Do They Exist?* Benjamins Translation Library 48. Benjamins, Amsterdam.

- MAY, RACHEL 1997: Sensible Elocution. How Translation Works in & upon Punctuation. – *The Translator* 3(1) s. 1–20.
- MINKKINEN, VIRPI 2001: Kirja 1980- ja 1990-luvuilla. Kirjallisuuden julkaiseminen, jakelu ja lukeminen. Kulttuuri ja viestintä. Tilastokeskus, Helsinki.
- MITCHELL, T. F. 1971: Linguistic ‘Goings On’: Collocations and Other Lexical Matters Arising on the Syntagmatic Record. – *Archivum Linguisticum* 2 s. 35–69.
- 1975: *Principles of Firthian Linguistics*. Longman, London.
- MÄKELÄ, MATTI 1993: Suomen murteiden *kin* ja *kaan, kään* -liitteet. Morfologia ja leksikko. SKS, Helsinki.
- MÄKIPÄÄ, MARJO 1995: Adjektiivin määritteet lehtikielessä. Pro gradu -tutkielma. Turun yliopisto.
- NELSON, MIKE 2000: A Corpus-Based Study of Business English and Business English Teaching Materials. Julkaisemanton väitöskirja. University of Manchester. Saatavilla: <http://www.kielikanava.com/thesis.html>. Luettu 1.8.2002.
- NEVALAINEN, SAMPO 2003: Käännöskirjallisuuden puhekielisyysistä – kaksinkertaista illuusiota? – *Virittäjä* 107 s. 2–26.
- (tulossa) Köyhtyykö kieli käännettäessä? Mitä frekvenssilistat kertovat suomennosten sanastosta. – Anna Mauranen – Jarmo H. Jantunen (toim.), *Käännösuomeksi: tutkimuksia suomennosten kielestä*. Tampere Studies in Language, Translation and Culture n:o 1.
- NS = *Nykysuomen sanakirja I–VI*. WSOY, Porvoo. 1951–61.
- OAKES, MICHAEL 1998: *Statistics for Corpus Linguistics*. Edinburgh University Press.
- OITTINEN, RIITTA 1995: *Kääntäjän karnevaali*. Tampere University Press.
- OJANSUU, HEIKKI 1909: Mikael Agricolan kielestä. *Suomi IV(7)*. SKS, Helsinki.
- OLKINUORA, KATRI 1992: Onko *aalto* suurempi kuin *laine*? Synonymian ongelmia. – Heikki Paunonen – Matti K. Suojanen – Kari Nahkola (toim.), *Variaatioita*. Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitoksen juhlakirja s. 113–124. Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
- OLOHAN, MAEVE – BAKER, MONA 2000: Reporting *that* in Translated English. Evidence for Subconscious Processes of Explicitation? – *Across Languages and Cultures* 1(2) s. 141–158.
- ONIKKI-RANTAJÄÄSKÖ, TIINA 2001: Sarjoja. Nykysuomen paikallissijaiset olotilanilmaukset kielen analogisuuden ilmentäjinä. SKS, Helsinki.
- ORPANA, TERTTU 1973: Intensiiteettiadvverbien semantiikkaa. *Painamaton lisensiaatintutkielma*. Tampereen yliopisto.
- 1988: Kuvaus vai kommentti. Tutkimus suomen kielen adjektiivadvverbien semanttisesta tulkinnasta. Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
- PAJUNEN, ANNELI 2000: Suomen ominaisuuspredikaatio ja tyyppi *olla sairaana*. – Anneli Pajunen (toim.), *Näkökulmia kielitypologiaan*. *Suomi* 186 s. 36–87. SKS, Helsinki.
- PALMER F. R. 1976: *Semantics. A New Outline*. Cambridge University Press.

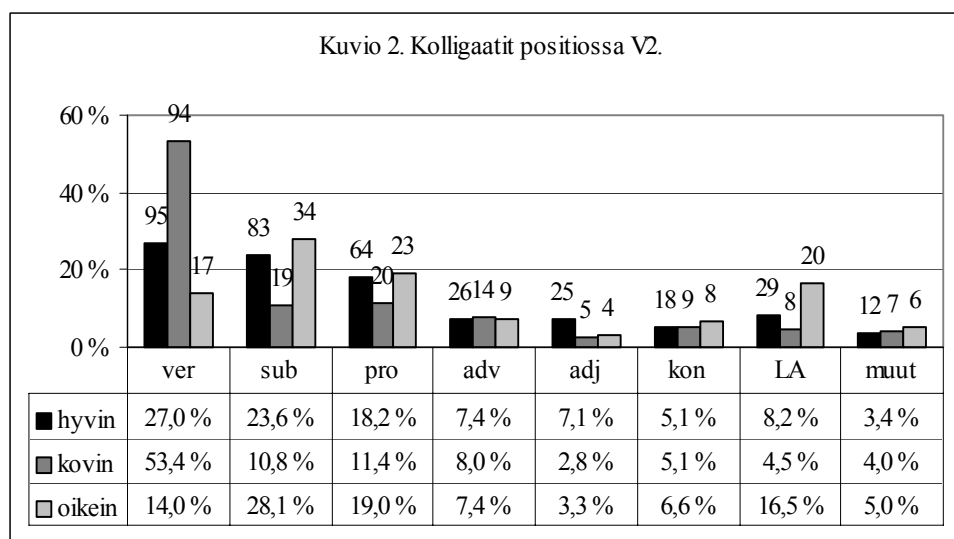
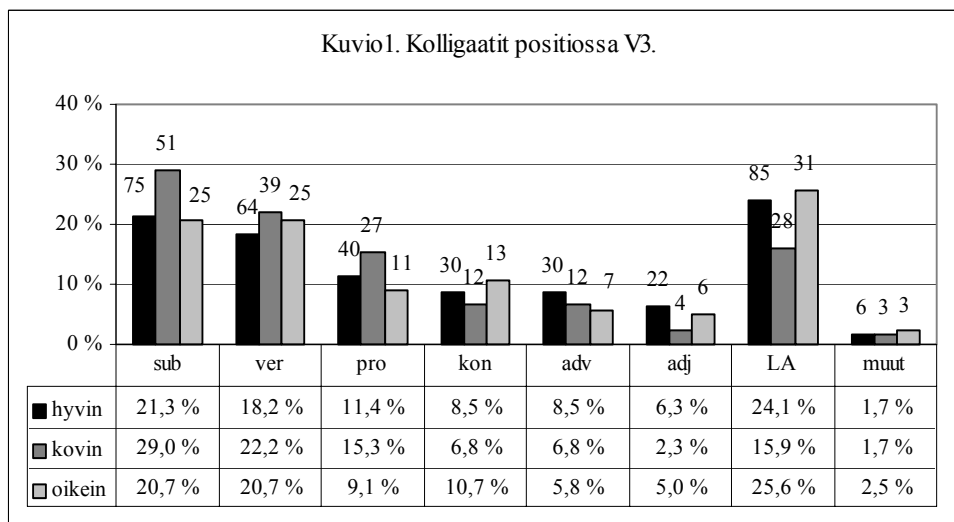
- PALOPOSKI, OUTI 2000: ”Nerontuotteita maailmankirjallisuudesta” – Suomenosten historia 1800-luvulta nykypäivään. – Outi Paloposki – Henna Makkonen-Craig (toim.), Käännöskirjallisuus ja sen kritiikki s. 20–33. AKO, Helsinki.
- 2001: Enriching translations, simplified language? An alternative viewpoint to lexical simplification. – *Target* 13(2) s. 265–288.
- 2002: *Variation in Translation. Literary Translation into Finnish 1809–1850*. Painamaton väitöskirja. Helsingin yliopisto.
- (tulossa) *Käännössuomen synty*. – Anna Mauranen – Jarmo H. Jantunen (toim.), *Käännössuomeksi: tutkimuksia suomennosten kielestä*. Tampere Studies in Language, Translation and Culture n:o 1.
- PARADIS, CARITA 1997: *Degree Modifiers of Adjectives in Spoken British English*. Lund University Press.
- PARTINGTON, ALAN 1998: *Patterns and Meanings. Using Corpora for English Language Research and Teaching*. Benjamins, Amsterdam.
- PEARSON, JENNIFER 1996: *Electronic texts and concordances in the translation classroom*. – *Teanga* 16 s. 85–95.
- PENTTILÄ, AARNI 1963: *Suomen kielioppi*. WSOY, Helsinki.
- PERSSON, GUNNAR 1989: *Deep and Profound: A Study in So-Called Synonymy*. Umeå Papers in English 10.
- 1990: *Meanings, Models and Metaphors. A Study in Lexical Semantics in English*. Almqvist & Wiksell International, Stockholm.
- PITKÄNEN, ANTTI – KOHONEN, VILJO 1984: *Johdatus kvantitatiiviseen kielentutkimukseen ja alan atk-sovelluksiin*. Gaudeamus, Helsinki.
- PUNTTILA, MATTI 1990: *Hyperhyvää vai huippuhuonoa? Tarpeellisista ja tarpeetomista vahvistussanoista*. – *Kielikello* 1/1990 s. 18–20.
- 1995: *Punsasta hyytelöön*. – Maija Länsimäki (toim.), *Sana kiertää: kirjoituksia suomen kielestä* s. 94–97. Kotimaisten kielten tutkimuskeskus, Helsinki.
- PUURTINEN, TIINA 1995: *Linguistic Acceptability in Translated Children’s Literature*. University of Joensuu Publications in the Humanities 15.
- Quirk ym. = QUIRK, RANDOLPH – GREENBAUM, SIDNEY – LEECH, GEOFFREY – SVARTVIK, JAN 1985: *A Comprehensive Grammar of the English Language*. Longman, London.
- RAPOLA, MARTTI 1960: *Sanojemme ensiesiintymiä Agricolasta Yrjö Koskiseen*. Tietolipas 22. SKS, Helsinki.
- 1967: *Suomenkielinen proosa ruotsin vallan aikana*. Tietolipas 49. SKS, Helsinki.
- REISS, KATHARINA 1971: *Möglichkeiten und Grenzen der Übersetzungskritik*. Max Hueber Verlag, Munich.
- – VERMEER HANS 1984: *Grundlegung einer allgemeinen Translations-theorie*. Niemeyer, Tübingen.
- RÄSÄNEN, MATTI 2002: *Harva se viikko*. – *Kielikello* 1/2002 s. 36.
- SADENIEMI, MATTI 1942: *Huomioita kielteisistä ja myönteisistä ilmauksista*. – *Virttäjä* 46 s. 285–289.

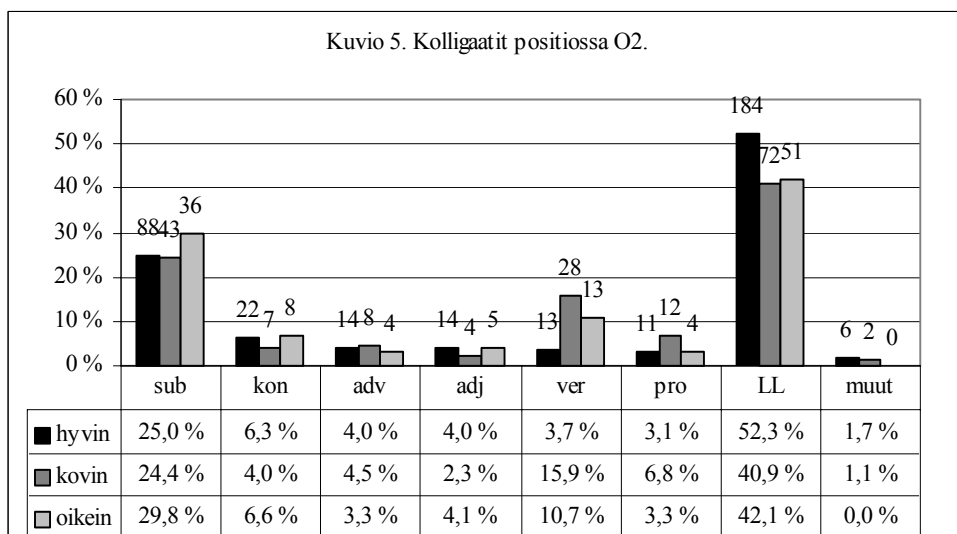
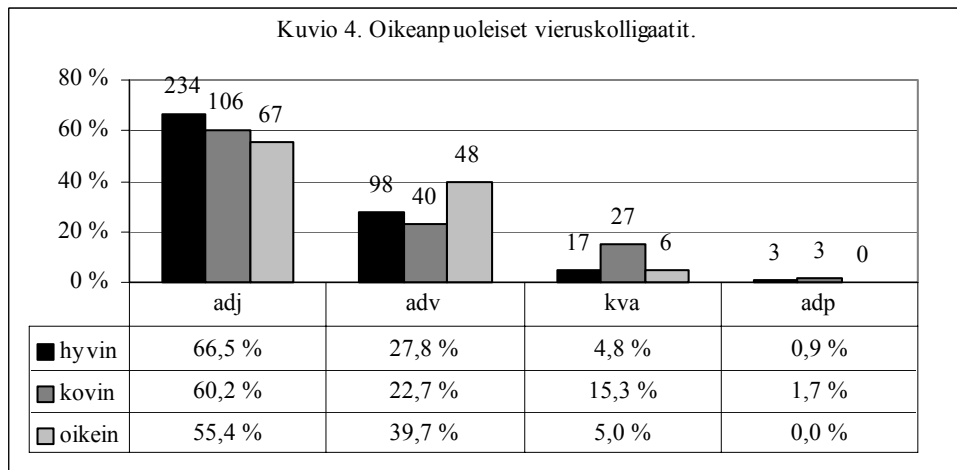
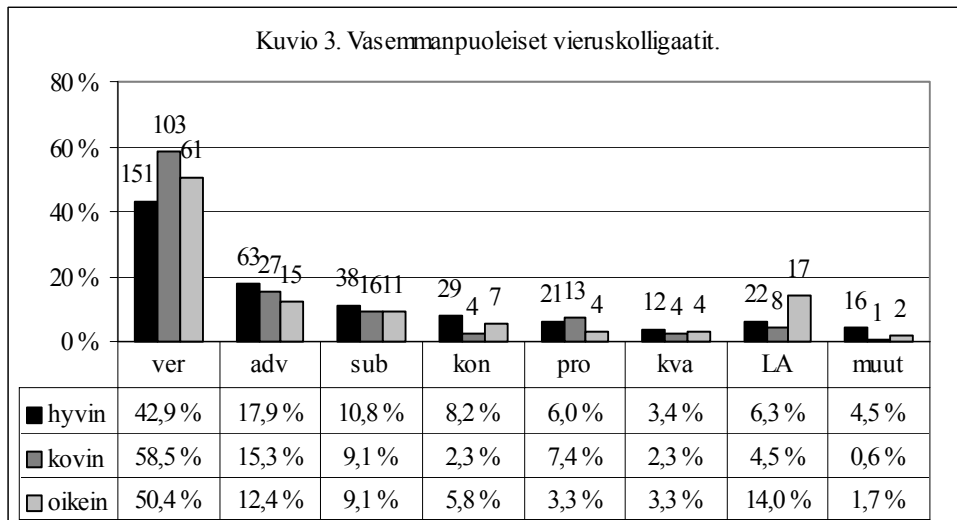
- SANDERS, WILLY 1977: *Linguistische Stilistik*. Vandenhoeck & Ruprecht, Göttingen.
- SAUKKONEN, PAULI 1984: *Mistä tyyli syntyy?* WSOY, Helsinki.
- – HAIPUS, MARJATTA – NIEMIKORPI, ANTERO – SULKALA, HELENA 1979: *Suomen kielen taajuussanasto*. WSOY, Helsinki.
- SAVIJÄRVI, ILKKA 1984: *Fakultaviivisten kieltoainesten ja kieltoverbin ellipsin suhteesta itämerensuomalaisissa kielissä*. – *Virittäjä* 88 s. 145–161.
- SCOTT, MIKE 1998: *WordSmith Tools manual, version 3.0*. Oxford University Press.
- SHLESINGER, MIRIAM 1991: *Interpreter Latitude vs. Due Process. Simultaneous and Consecutive Interpretation in Multilingual Trials*. – Sonja Tirkkonen-Condit (toim.), *Empirical Research in Translation and Intercultural Studies. Selected Papers of the TRANSIF Seminar, Savonlinna 1988* s. 147–155. Gunter Narr, Tübingen.
- SHUTTLEWORTH, MARK – COWIE, MOIRA 1997: *Dictionary of Translation Studies*. St. Jerome, Manchester.
- SINCLAIR, JOHN 1966: *Beginning the Study of Lexis*. – C. Bazell – J. Catford – M. A. K. Halliday – R. Robins (toim.), *In Memory of J. R. Firth* s. 410–430. Longman, London.
- 1987: *Collocation: a progress report*. – Ross Steele – Terry Threadgold (toim.), *Language Topics. Essays in Honour of Michael Halliday 2* s. 319–331. Benjamins, Amsterdam.
- 1991: *Corpus, Concordance, Collocation*. Oxford University Press.
- 1996: *The Search for Units of Meaning*. – *Textus IX* s. 75–106.
- 1998: *The Lexical Item*. – Edda Weigand (toim.), *Contrastive Lexical Semantics* s. 1–24. Benjamins, Amsterdam.
- 2003: *Reading concordances. An Introduction*. Longman, London.
- SKS = Suomalaisen Kirjallisuuden Seura. Helsinki.
- SMADJA, FRANK 1991: *Macrocoding the Lexicon with Co-occurrence Knowledge*. – Uri Zernik (toim.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon* s. 165–189. Lawrence Erlbaum, London.
- 1993: *Retrieving Collocations from Text: Xtract*. – *Computational Linguistics*. 19(1) s. 143–177.
- STASSEN, LEON 1997: *Intransitive predication*. Clarendon Press, Oxford.
- STEWART, DOMINIC 2000: *Conventionality, Creativity and Translated Text: The Implications of Electronic Corpora in Translation*. – Maeve Olohan (toim.), *Intercultural faultlines. Research models in translation studies I. Textual and cognitive aspects* s. 73–91. St. Jerome, Manchester.
- STUBBS, MICHAEL 1995a: *Collocations and semantic profiles. On the Cause of the Trouble with Quantitative Studies*. – *Functions of Language 2*(1) s. 23–55.
- 1995b: *Corpus evidence for norms of lexical collocation*. – Guy Cook – Barbara Seidlhofer (toim.), *Principle and Practice in Applied Linguistics* s. 245–256. Oxford University Press.
- 1996: *Text and Corpus Analysis. Corpus-assisted Studies of Language and Culture*. Blackwell, Oxford.

- 2001a: *Words and Phrases. Corpus Studies of Lexical Semantics*. Blackwell, Oxford.
- 2001b: On inference theories and code theories: Corpus evidence for semantic schemas. – *Text* 21(3) s. 437–465.
- TEUBERT, WOLFGANG 1996: Comparable or Parallel Corpora? – *International Journal of Lexicography* 9(3) s. 238–264.
- TIRKKONEN-CONDIT, SONJA 2000a: Kääntämisen teoria, tutkimus ja sovellus. – Kari Sajavaara – Arja Piirainen-Marsh (toim.), *Näkökulmia soveltavaan kielentutkimukseen* s. 123–154. Soveltavan kielentutkimuksen keskus, Jyväskylä.
- 2000b: In Search of Translation Universals: Non-equivalence of 'Unique' Items in a Corpus Test. Esitelmä konferenssissa *Research Models in Translation Studies* 28–30 April 2000. Manchester, UK.
- 2001: Unique Items – Over- or Under-represented in Translated Language? Esitelmä konferenssissa *Translation Universals – Do They Exist?*, Savonlinna, 19.–20.10.2001.
- 2002: Translationese; a Myth or an Empirical Fact? A Study into the Linguistic Identifiability of Translated Language. – *Target* 14(2) s. 207–220.
- 2004: Unique items – over or under-represented in translated language? – Anna Mauranen – Pekka Kujamäki (toim.), *Translation Universals. Do They Exist?* s. 177–184.
- (tulossa) Häviävätkö uniikkiaineokset käännösuomesta? – Anna Mauranen – Jarmo H. Jantunen (toim.), *Käännösuomeksi: tutkimuksia suomennosten kielestä*. Tampere Studies in Language, Translation and Culture n:o 1.
- TOGNINI-BONELLI, ELENA 1996: *Corpus Theory and Practice*. TWC, Birmingham.
- 2001: *Corpus Linguistics at Work*. Benjamins, Amsterdam.
- TOURY, GIDEON 1974: *Li-she'elat te'ur ha-sifrut ke-rav ma'arekhet* [Literature as a Polysystem]. – *Ha-sifrut/Literature* 18–19 s. 1–19.
- 1980: *In Search of A Theory of Translation*. The Porter institute for Poetics and Semiotics, Tel Aviv University.
- 1985: A Rationale for Descriptive Translation Studies. – Theo Hermans (toim.), *Manipulation of Literature. Studies in Literary Translation* s. 16–41. Groom Helm, London.
- 1991: What are Descriptive Studies into Translation Likely to Yield apart from Isolated Descriptions? – Kitty van Leuven-Zwart – Ton Naaijken (toim.), *Translation Studies: The State of the Art. Proceedings of the First James S. Holmes Symposium on Translation Studies* s. 179–192. Rodopi, Amsterdam.
- 1995: *Descriptive Translation Studies and beyond*. Benjamins, Amsterdam.
- 2004: Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals? – Anna Mauranen – Pekka Kujamäki (toim.), *Translation Universals. Do They Exist?* s. 15–32.
- TUOMIKOSKI, RISTO 1969: *Asemoisilmaustemme luokittelusta*. Suomi 114(4). SKS, Helsinki.

- 1973: Adjektiivin ja adverbien erosta. – *Virittäjä* 77 s. 199–213.
- 1978: Objektinsijaisista määrän adverbiaaleista. – *Virittäjä* 82 s. 26–51.
- TYMOCZKO, MARIA 1998: Computerized Corpora and the Future of Translation Studies. – *Meta* 43(4) s. 652–659.
- VANDERAUWERA, RIA 1985: Dutch Novels Translated into English: The Transformation of a “Minority” Literature. Rodopi, Amsterdam.
- VANHATALO, ULLA 2003: Kyselytestit vs. korpuslingvistiikka lähisynonyymien semanttisten sisältöjen arvioinnissa – Mitä vielä *keskeisestä* ja *tärkeästä*? – *Virittäjä* 107 s. 351–369.
- WARREN, BEATRICE 1988: Semantics: Word meaning. – Nils-Lennart Johannesson (toim.), *The Dress of Thought. Aspects of the Study of Language* s. 61–95. Studentlitteratur, Lund.
- VEHMAS-LEHTO, INKERI 1989: Quasi-correctness. A critical study of Finnish translations of Russian journalistic texts. Neuvostoliittoinstituutin vuosikirja 31. Helsinki.
- 1999: Kopiointia vai kommunikointia. Johdatus käännteoriaan. Finn Lectura, Helsinki.
- WETZER, HARRIE 1996: *The Typology of Adjectival Predication*. Mouton de Gruyter, Berlin.
- VILKUNA, MARIA 1996: Suomen lauseopin perusteet. Kotimaisten kielten tutkimuskeskus, Helsinki.
- VILPPULA, MATTI 2001: Erinomaisen ilkeä ihminen. *Kieli-ikkuna*. HS 26.6.2001.
- ZANETTIN, FEDERICO 2000: Parallel Corpora in Translation Studies: Issues in Corpus Design and Analysis. – Maeve Olohan (toim.), *Intercultural Faultlines. Research Models in Translations Studies* 1. Textual and cognitive Aspects 105–118. St. Jerome, Manchester.

LIITE 1. Astemääritteiden kolligaattien osuudet AS-aineistossa.





LIITE 2. O1-kollokaattien ryhmittely semanttista preferenssiä kuvaaviin luokkiin.

Kollokaatit on lueteltu taulukkoon lemminoina.

| | <i>hyvin</i> | <i>kovin</i> | <i>oikein</i> |
|---------------------------------|--|--|---|
| aika / ikä | aikaisin, harvoin, hidas, hitaasti, kauan, kaukainen, nopeasti, nuori, pian, pitkään, satunnainen, säännöllinen, tiuha, usein, vanha, äkillisesti, äkkiä | kauan, myöhään, pitkään, sukkelasti, usein, vanha | pitkään, vanha |
| arvo | aito, huonosti, hyvä, juhlaallinen, kallisarvoinen, kallis, käyttökelpoinen, merkittävä, myönteisesti, nihkeästi, tähdellinen, tärkeä, vakava-rainen, varakas, varteenotettava | hauras, hedelmällinen, huono, hyvin, hyvä, korkea, kummoinen, lahjakas, merkillinen, onnistunut, paha, tärkeä, vaatimaton, varakas | hyvin, hyvä, kurja, mainio, onnistunut, paha |
| fyysinen tila / aistimus | elävä, eroottinen, herkkä, hiljaa, hiljainen, houkuttava, kevyt, kiihottava, kuol-lut, kuuma, kylmä, lämmin, maallishukumallinen, olfaktorinen, pehmeä, ruumiillinen, rytmikkäästi, räyhäkäs, seksuaalinen, väsynyt, äänekkäästi | elävä, haluttava, lämmin, maskuliininen, raskas, äänekkäästi | hyväkuntoinen, kireä, kuuluva |
| intensi-teetti | intensiivisesti, kovaa, pää-määrätietoisesti, raa'asti, rajullinen, rauhallisesti, tarkkaan, tehokas, vahva, vahvasti, vakava, varovasti, varuillaan, voimakas, voimakkaasti, väkevästi | lujaa, painokkaasti, ponteva, rauhallisesti, syvä, syvällisesti, tiukassa, vakavasti, voimakas | harras, kovasti, pahasti, perusteellisesti, raskas, roimasti, sitkeästi, syvästi, vakavasti, voimakkaasti |
| kognitiivisuus | epäilevä, epäluuloinen, kiinnostunut, kriittisesti, kunnioitettava, kysyvästi, mielenkiintoinen, määrätietoinen, nerokas, opettavainen, paikkansapitävä, ristiriitaisesti, utelias, uteliaasti, ymmärrettävä | harkittu, johdonmukainen, järkevä, kiinnostava, kiinnostunut, käytännöllinen, älykäs | kokenut, viisas |
| luonne | arka, avomielisesti, hellä, hellästi, humaani, huomaa-vainen, hyökkäävästi, itse-käs, itsenäinen, karkeasti, kohtelias, viehättävä, vilkas, yleiskiltisti ystävällinen, ystävällisesti | ahne, hauska, kirjallinen, luokkatiетoinen, miehinen, röyhkeä, utelias, vanhanai-kainen, viaton, ylevä, ystä-vällinen | hauska, mellevä, sopuisa, valoisasti |
| määrä | harva, hyvin, moni, paljon, rajoitetussa määrin, sääste-liäästi, vähän, yksin | moni, niukasti, paljon, suuresti, tuottoisa, täydelli-sesti, vilkas, vähän, yksin | kovasti, moni, paljon, tuke-va |
| paikka | bysanttilainen, kauas kaukainen, korkealla, lähel-lä, ylhäällä | kauas, lähellä | lähellä, ylhäällä |

| | | | |
|---------------------------------------|--|---|---|
| tarkkuus / selvyys | huolellisesti, hämärä, hämä-
rästi, kirrkaasti, konkreetti-
sesti, salaperäinen, selkeä,
selvä, selvästi, tarkka,
tarkkaan, tarkkasilmäinen,
tarkoin | epämääräinen, huolellisesti,
tarkasti, tarkkaan | selvästi, tarkka, tarkasti,
yksityiskohtaisesti |
| tunne | epämiellyttävä, hermostu-
nut, hämmentynyt hämmäs-
tynyt, ilkeä, ilkeästi, iloi-
nen, inhottava, innokas,
innostunut, julmasti, kiuk-
kuinen, kiusallinen, kolikko,
kunnioitettava, kylmäkis-
koisesti, lempeästi, levoton,
lohdullinen, loukkaava,
miellyttävä, mukava, mus-
tasukkainen, onnellinen,
pahoillaan, pettynyt, rakas-
tettava, rauhallinen, raskas,
soma, surullinen, tyyni,
tyytymätön, vastenmieli-
nen, viehättävä, yksinäinen,
yllätynyt | haltioituneesti, herkkä,
iloinen, innostunut, kiih-
dyksissä, kiva, liikuttava,
läheinen, mielellään, nolo,
onneton, palkitsevaa, petty-
nyt, surullinen, vastenmie-
linen, ärtynyt | hurmaava, ihailtava, ihanas-
ti, ikävä, kiva, mielellään,
mieluisuinen, mukava, petty-
nyt, tunnelmallinen, tyyty-
väinen, vihainen |
| tuttuus / erilaisuus | arkisesti, epätodellinen,
erikoinen, erikoislaatuinen,
henkilökohtainen, kummal-
linen, outo, realistinen,
salainen, sovinnainen,
suosittu, tavallinen, toisen-
lainen, tuttu, tyypillinen,
vaihteleva, yksilöllinen,
yleinen, yllättävä | erikoinen, erilainen, kum-
mallinen, outo, sama, taval-
linen, todellinen, tuttu | – |
| ulkonäkö | elegantti, erinäköinen,
hento, kaunis, paljastava,
puhdas, punainen, raffinoi-
tu, sievä, sininen, sinisil-
mäinen, suloinen, tumma,
vaalea, valkoinen, valoisa | edustava, hentorakenteinen,
huterannäköinen, kalpea,
tummahiipiäinen, tyylilik-
käästi | kaunis, rintava, sievä, soma,
viehkeä, viehättävä |
| ulottuvuus | eripituinen, kookas, korkea,
laiha, lyhyt, matala, matala-
jalkainen, pieni, pitkä,
suppea, suuri, syvä | ahdas, iso, kookas, lyhyt,
pieni, pitkä, rajallinen, suuri | iso, pieni, suuri |
| vaikeus | mutkallinen, ongelmallinen,
vaikea, yksinkertainen | hankala, helposti, helppo,
luontevasti, vaikea, yksin-
kertainen | mahdoton, vaikea |
| varmuus | epätodennäköinen, mahdol-
linen, todennäköinen, to-
dennäköisesti | todennäköinen | – |
| luokitte-
lematto-
mat | käytännöllinen, luvaton,
piikkioksaainen, sateinen,
syvä, takertuva, uskonnolli-
nen, vaativa | musikaalinen | – |

| | | | | | | | | | | | | | | |
|-----------|---|-----|-----|------|----------------|----|-----|------|------|-----------------------|----|-----|------|------|
| sitten | 7 | 0,4 | 586 | 33,3 | aina | 12 | 0,3 | 1088 | 30,7 | paljon yksinkertainen | 10 | 0,4 | 970 | 35,0 |
| vanha | 7 | 0,4 | 593 | 33,7 | itse | 12 | 0,3 | 1100 | 31,0 | tämä | 10 | 0,4 | 980 | 35,3 |
| väsinyt | 7 | 0,4 | 600 | 34,1 | tämä | 13 | 0,4 | 1113 | 31,4 | ei | 10 | 0,4 | 990 | 35,7 |
| katsoa | 6 | 0,3 | 606 | 34,4 | vaikkea | 12 | 0,3 | 1125 | 31,7 | nähdä | 9 | 0,3 | 999 | 36,0 |
| kysyä | 6 | 0,3 | 612 | 34,8 | paljon suuri | 11 | 0,3 | 1136 | 32,0 | onnellinen | 9 | 0,3 | 1008 | 36,3 |
| nainen | 6 | 0,3 | 618 | 35,1 | tai | 11 | 0,3 | 1147 | 32,4 | pian vähän | 9 | 0,3 | 1017 | 36,6 |
| nyt | 6 | 0,3 | 624 | 35,5 | näyttää | 11 | 0,3 | 1158 | 32,7 | asia | 9 | 0,3 | 1026 | 37,0 |
| näyttää | 6 | 0,3 | 630 | 35,8 | pian vain | 10 | 0,3 | 1168 | 32,9 | hitaasti | 9 | 0,3 | 1035 | 37,3 |
| pian | 6 | 0,3 | 636 | 36,1 | asia | 10 | 0,3 | 1178 | 33,2 | kummallinen | 8 | 0,3 | 1043 | 37,6 |
| alkaa | 5 | 0,3 | 641 | 36,4 | elämä | 10 | 0,3 | 1188 | 33,5 | mikä | 8 | 0,3 | 1051 | 37,9 |
| kaunis | 5 | 0,3 | 646 | 36,7 | ihminen | 9 | 0,3 | 1197 | 33,8 | paha | 8 | 0,3 | 1059 | 38,2 |
| moni | 5 | 0,3 | 651 | 37,0 | kaus | 9 | 0,3 | 1206 | 34,0 | tehdä | 8 | 0,3 | 1067 | 38,5 |
| nuori | 5 | 0,3 | 656 | 37,3 | kaunis | 9 | 0,3 | 1215 | 34,3 | vaikkea | 8 | 0,3 | 1075 | 38,7 |
| paljon | 5 | 0,3 | 661 | 37,6 | ottaa | 9 | 0,3 | 1224 | 34,5 | kaikki | 8 | 0,3 | 1083 | 39,0 |
| vaikka | 5 | 0,3 | 666 | 37,8 | puhua | 9 | 0,3 | 1233 | 34,8 | kylmä | 8 | 0,3 | 1091 | 39,3 |
| varovasti | 5 | 0,3 | 671 | 38,1 | selvästi | 9 | 0,3 | 1242 | 35,0 | moni | 8 | 0,3 | 1099 | 39,6 |
| voida | 5 | 0,3 | 676 | 38,4 | sitten (adv.) | 9 | 0,3 | 1251 | 35,3 | muu | 7 | 0,3 | 1106 | 39,9 |
| | | | | | tapa | 8 | 0,2 | 1260 | 35,5 | poika | 7 | 0,3 | 1113 | 40,1 |
| | | | | | tehdä | 9 | 0,3 | 1268 | 35,8 | sillä | 7 | 0,3 | 1120 | 40,4 |
| | | | | | harvoin | 9 | 0,3 | 1277 | 36,0 | silti | 7 | 0,3 | 1127 | 40,6 |
| | | | | | lähellä (adv.) | 8 | 0,2 | 1286 | 36,3 | suuri | 7 | 0,3 | 1134 | 40,9 |
| | | | | | myös | 8 | 0,2 | 1294 | 36,5 | tiestä | 7 | 0,3 | 1141 | 41,1 |
| | | | | | nainen | 8 | 0,2 | 1302 | 36,7 | väsinyt | 7 | 0,3 | 1148 | 41,4 |
| | | | | | vaarallinen | 8 | 0,2 | 1310 | 37,0 | aina | 7 | 0,3 | 1155 | 41,6 |
| | | | | | yksinkertainen | 8 | 0,2 | 1318 | 37,2 | koska | 7 | 0,3 | 1162 | 41,9 |
| | | | | | erikoinen | 8 | 0,2 | 1326 | 37,4 | kuollut | 7 | 0,3 | 1169 | 42,1 |
| | | | | | hitaasti | 8 | 0,2 | 1334 | 37,6 | niin | 6 | 0,2 | 1175 | 42,3 |
| | | | | | jo | 7 | 0,2 | 1341 | 37,8 | pimeä | 6 | 0,2 | 1181 | 42,6 |
| | | | | | kaksi | 7 | 0,2 | 1348 | 38,0 | saada | 6 | 0,2 | 1187 | 42,8 |
| | | | | | kyllä | 7 | 0,2 | 1355 | 38,2 | | 6 | 0,2 | 1193 | 43,0 |
| | | | | | | 7 | 0,2 | 1362 | 38,4 | | 6 | 0,2 | 1199 | 43,2 |
| | | | | | | 7 | 0,2 | 1369 | 38,6 | | 6 | 0,2 | 1205 | 43,4 |

Jarmo Harri Jantunen

Synonymia ja käännössiuna: korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin Joensuun yliopiston humanistisia julkaisuja 35, 2004

Errata

| | | |
|--|--------------------------------|---|
| s. 5 | luku 1.3, 1. kappale, rivi 13 | elektronis <u>i</u> a kieliaineistoja |
| s. 5 | luku 1.4. rivi 3 | aiemmassa tutkimuk <u>s</u> essa |
| s. 14 | 2. kappale, rivi 7 | näennäisenä, sill <u>ä</u> |
| s. 31 | 2. kappale, rivi 11 | (19 <u>6</u> 6: 429) |
| s. 34 | luku 3.1.1, 1. kappale, rivi 8 | dynaamist <u>e</u> n |
| s. 52 | 2. kappale, rivi 1 | kielen yksin <u>n</u> kertaistumisesta |
| s. 65 | luku 5.1 kohta 2.2 | valinta <u>e</u> i ole yksinkertaistunutta |
| s. 109 | 2. kappale, rivi 6 | konkordanssissa 6 <u>9</u> rivit |
| s. 111 | toiseksi viimeinen rivi | tai toivot <u>u</u> sta, jotka ovat |
| s. 112 | konkordanssi 6.11 | |
| <p>4 oli kiva seurata LA miten <i>oikein</i> vanha <u>i</u>a kömpelö mies
5 niin, että meidän murheemme ovat <i>kovin</i> pieniä <u>i</u>a ohimeneviä.</p> | | |
| s. 123 | rivi 6 | huomattavasti <u>m</u> m. muita pitempi |
| s. 125 | rivi 9 | taulukkoa 6.7 s._98) |
| s. 133 | 1. kappale, rivi 6 | ennalta-arvattavuutta |
| s. 153 | kolmanneksi alin rivi | yliedustettuja K <u>S</u> E-aineistossakin |
| s. 158 | rivi 3 | aineistossa vain <u>2</u> kaksi esiintymää |
| s. 175 | taulukko 7.10 | semanttinen prosodia: negatiiv <u>i</u> nen
prosodia, positiiv <u>i</u> nen prosodia |
| s. 177 | 2. kappale, rivi 11 | modaalisuudesta_Ahlman 1933 |
| s. 183 | 2. kappale, rivi 11 | ei ole V1-positioissa ainuttakaan |
| s. 191 | 1. kappale, rivi 3 | positioihin <u>V</u> 2 ja <u>V</u> 3. Tällaisia |
| s. 208 | 2. kappale, rivi 2 | klusteroitumisasteiden <u>e</u> rojen
selvänä |
| s. 219 | 2. kappale, viim. rivi | merkkejä kote <u>k</u> stuaalisten
piirteiden |