

Kielten osasto
Humanistinen tiedekunta
Helsingin yliopisto

KURKISTUKSIA KOLLOKAATIOON

EU-suomen *toimenpiteen* rantautumisesta
Suomen säädöskieleen

Hanna Westerlund

Akateeminen väitöskirja, joka Helsingin yliopiston humanistisen
tiedekunnan suostumuksella esitetään julkisesti tarkastettavaksi
Porthanian salissa P674 26. syyskuuta 2020 klo 12.

Helsinki 2020

Ohjaajat	Lauri Carlson, Helsingin yliopisto Jussi Piitulainen, Helsingin yliopisto
Esitarkastajat	Antti Arppe, University of Alberta Jarmo Harri Jantunen, Jyväskylän yliopisto
Vastaväittäjä	Jarmo Harri Jantunen, Jyväskylän yliopisto

© Hanna Westerlund

ISBN 978-951-51-6564-0 (nid.)
ISBN 978-951-51-6565-7 (PDF)

Unigrafia
Helsinki 2020

Leon 

Kurkistuksia kollokaatioon: EU-suomen *toimenpiteen* rantautumisesta Suomen säädöskieleen

Hanna Westerlund

hanna.westerlund@iki.fi

Abstract

In translation studies, the notion of collocation has gained growing interest during the past 20 years. However, no generally accepted definition of the concept is as of yet available. Collocation is rather defined from various points of view, subject to the domain and research interests. As part of my research project, I seek to form a coherent picture of collocation, and to observe properties of collocation in order to identify instances of collocation in corpora. Rooted in the Firthian tradition of language studies, the theory is reflected upon the findings which stem from the data observed. The properties identified comprise frequency, structure, meaning and association, with the focus clearly on the first two.

The present research seeks to find out whether collocations in the translated European Union regulations are similar to those in statutes originally written in Finnish. For this purpose, two corpora were created, and three reference corpora provided by the Language Bank of Finland were consulted during the research process. The research corpus comprises regulations by the European Union at the time Finland joined the community, and the first reference corpus comprises parts of the Finnish legislation from the same period. The findings in these two corpora are compared with the Semfinlex corpora of original statutes, the corpus of laws and directives originally compiled at the Institute for the Languages of Finland, and a corpus of newspapers and magazines. The research methods are basic language technology tools, and the method is descriptive in nature.

One concurrent term is studied in detail, *toimenpide* ['measure'], and its collocative behaviour with six verbs is recorded. These verbs are *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä* and *toteuttaa* [verbs of 'doing']. Besides offering insights into theory formation, the results show a clear tendency for the translations of the EU regulations to display verbs differing from the Finnish legislation to form collocations with the term subject to research. Analysis of the collocates in various corpora suggests that there is a change taking place in the Finnish legislative language with more variation in the usage of the term and new meanings added to the existing ones.

Tiivistelmä

Kollokaatio on ollut käännöstieteen tutkijoiden kasvavan kiinnostuksen kohteena viimeisen 20 vuoden ajan. Ilmiön määrittelyssä ei ole saavutettu yksimielisyyttä ja kollokaatio tutkimuskohteena määritellään tavallisesti tutkimuskohteeseen ja tutkimustavoitteisiin nähden tarpeellisella laajuudella. Olen koonnut mahdollisimman kattavan joukon tutkimuskirjallisuudesta löytyviä ominaispiirteitä, joiden avulla arvioin ja käsitelen aineistosta nousevia havaintoja. Ominaispiirteet ovat määrällisiä, rakenteeseen liittyviä, merkitystä ja assosiaatiota valottavia. Näistä määrälliset ja kollokaation rakenteeseen liittyvät ominaispiirteet ovat tutkimukseni keskiössä. Tarkastelen merkitykseen ja assosiaatioon liittyviä piirteitä siinä määrin, kun se aineiston ja tutkimusmenetelmän ehdoilla on mahdollista.

Näkökulmani tutkimuksessa on kääntäjän ja katson käännettävään tekstiin pyrkien näkemään sen termien tendenssiä kollokoida muiden kielellisten elementtien kanssa. Pyrin selvittämään, ovatko Euroopan unionin suomennetut asetukset kollokaatioiden suhteen erilaisia kuin Suomen säädöskokoelman säädökset ja saapuuko EU-asetusten myötä Suomen säädöskieleen epätyypillisiä kollokaatioita. Varsinaisen tutkimusaineistoni koostuu EU-asetuksista ajalta, jolloin Suomi liittyi Euroopan yhteisöjen jäseneksi. EU-asetusten ensimmäiseksi verrannaisaineistoksi olen koonnut Suomen säädöskokoelmasta sisällöllisesti ja ajallisesti mahdollisimman hyvin varsinaista tutkimusaineistoa vastaavan tekstikorpuksen. Näistä kahdesta aineistosta tehtyjä havaintoja ja tuloksia vertaan Kielipankin kautta käytettävissä olevaan Semanttisen finlexin lakeja sisältävään osakorpukseen, Kotuksen lakeja, asetuksia ja direktiivejä sisältävään tekstikorpukseen sekä yleiskieliseen sanoma- ja aikakauslehtiä sisältävään tekstikorpukseen. Tutkimusote on deskriptiivinen ja käytän tutkimusmenetelminä kieliteknologian kehittämiä avoimia työkaluja.

Keskityn tarkastelemaan termiä *toimenpide* ja sen yhteisesiintymistä tekstiyhteydessä esiintyvän kuuden verbin kanssa. Nämä verbit ovat *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä* ja *toteuttaa*. Tulokset osoittavat selkeitä määrällisiä eroja EU-aineiston ja Suomen säädöskokoelman välillä aineistoissa havaittavien yhteisesiintymien suhteen. EU näyttäytyy uusia ajattelutapoja ja uutta sääntelyä suomalaisen säädöskielen tuovana, mutta myös uudenlaisia merkityssisältöjä tutkimuskohde *toimenpiteelle* tuottavana organisaationa. Tulosten perusteella voidaan olettaa, että suomalainen säädöskieli tulee muuttumaan kohdatessaan Euroopan unionin tekstejä. Suomen säädöskielessä siirrymme *toteuttamaan toimenpiteitä* niihin *ryhtymisen* sijaan.

Kiitokset

Ajattelu on haastavaa. Selkeältä vaikuttava asia muuttuu sitä monesta suunnasta pohtiessa epätarkaksi ja sitä kautta haastavaksi. Opitun uudelleen rakentaminen tuntuu vaikealta. Tutkimusmatkani varrella olen kohdannut alati vaihtoehtoisia tapoja mieltää, rajata ja ymmärtää tutkimuskohdettani paljon useammin kuin olisin toivonut. Tässä tulos nyt kuitenkin on, havainnot aineiston suomennosten yhden termin esiintymistä rajatussa tekstikontekstissa. Muitakin kiinnostavia termejä nousi aineistosta esiin, mutta lopulta tutkimus palasi takaisin lähtöpisteeseen: termiin, joka heti alkumetreillä nousi esiin vuosituhaten vaihteessa silloisella Joensuu yliopiston kansainvälisen viestinnän laitoksella Savonlinnassa tulostamassani avainsanaluettelossa.

Tutkimuksen vieminen loppuun on ollut mahdollista ainoastaan lukemattomien ihmisten tuella. Sen käynnistävänä ja kaiken aikaa kantavana voimana on ollut ohjaajani Lauri Carlson, joka on kannustanut jatkamaan niinä lukuisina kertoina, jolloin olen ollut valmis unohtamaan koko asian. Lämmin kiitos, Lauri. Ohjaajani Jussi Piitulainen on väsymättä kuunnellut pohdintoja tilastollisista menetelmistä. Ilman Jussia työ ei olisi valmistunut. Kiitos, Jussi. Kiitän myös lämpimästi esitarkastajiani, Jarmo Harri Jantusta Jyväskylän yliopistosta ja Antti Arppea Albertan yliopistosta kärsivällisyydestä, jaksamisesta ja aivan ratkaisevan tärkeistä kommentteista. Kommenttien myötä työ on lopulta tullut päättepisteeseensä.

Matkallani minulla on ollut ilo saada olla mukana lukuisissa tutkimusyhteisöissä, joista kaikki ovat tarjonneet tarpeellisia rakennuspalikoita. Savonlinnan kansainvälisen viestinnän laitoksen väen, erityisesti Anna Maurasen urauurtava työ käännösuomen aineistojen kokoamisessa oli ratkaisevan tärkeää ja hänen johtamansa universaalien tutkimuksen tutkimusryhmä ja sen tohtorikoulutettavat Sari Eskola, Jarmo Harri Jantunen ja Hannu Kemppanen tarjosivat oivan lähtökohdan työn alkumetreillä. Tutkijakoulu Langnet on monellakin tapaa merkityksellinen yhteisö, joka on pitänyt tutkimuksen mielessäni muun elämän huiskeessa. Erityisen kiitollinen olen Liisa Tiittulalle ja Andrew Chestermanille kaikesta tuesta kiperissäkin paikoissa. Sydämellinen kiitos, Liisa ja Andrew. Kiitos, Langentin käännöstieteen tutkijat kautta aikojen hienoista seminaareista, illanvietoista ja teatteriesityksistä lukuisissa karonkoissa. On ollut korvaamattoman hienoa saada olla niissä mukana. Kiitos Outi Paloposki, Leena Salmi, Kaisa Koskinen, Tuija Kinnunen, Henna Makkonen-Craig ja monet muut.

Helsingin yliopiston yleisen kielitieteen laitoksella kokoontui 2000-luvun puolivälissä upea TR9-ryhmä, jonka jäsenten esitelmistä harvemmin ymmärsin sisältöjä, mutta jotka auttoivat hahmottamaan kielitieteellisen tutkimuksen kenttää. Kiitos Anu Airola, Antti Arppe, Urho Määttä, Juhani Järvikivi, Camilla Magnusson, Jussi

Piitulainen, Reetta Konstenius ja Martti Vainio. Kiitos yleisen kielitieteen laitoksen väelle ja Kimmo Koskenniemelle ja Fred Karlssonille mahtavasta työympäristöstä ja Johanna Ratialle ja Johanna Hirvoselle sekä lukuisille kieliteknologian opiskelijoille hienoista vuosista Siltavuorenpenkereellä. Kouvolan kääntäjänkoulutuslaitoksella pääsin pohtimaan käännöstieteellisiä näkökulmia, mistä kiitos Irmeli Helinille, Hilka Yli-Jokipiille, Inkeri Vehmas-Thesslundille ja Ritva Hartama-Heinoselle. Kiitos Helsingin yliopiston tutkijayhteisö TRASTin jäsenille mielenkiintoisista keskusteluista seminaareissa ja niiden ulkopuolella, erityisesti Mari Pakkala-Weckströmille ja Päivi Pasaselle. Kiitos ystäväilleni Erja Tenhonen-Lightfootille, Stina Ojalalle ja Henna Makkonen-Craigille ja monille muille matkan varrella kohtaamilleni hienoille ihmisille kaikesta avusta ja kannustuksesta. Kaunis kiitos Kielipankin tiimille Kristter Lindén, Mietta Lennes, Jussi Piitulainen, Jyrki Niemi, Erik Axelson, Ute Dieckmann, Tero Aalto ja Martin Matthiesen tuesta vaikeina vuosina sekä kielentutkijoille niin tärkeiden aineistojen tuottamisesta. Kiitos Martti Vainiolle ja Stefan Wernerille luottamuksesta. Kiitos kieliteknologi Anssi Yli-Jyrälle aina niin positiivisesta suhtautumisesta ja kielitieteilijä Kaius Sinnemäelle rohkaisevista keskusteluista. Kiitos myös Teatterikorkeakoulun väelle, erityisesti Annika Fredrikssonille ja Elina Raitasalolle henkisestä tuesta kautta aikojen. Kiitos Katariina Lehtolalle kannustamisesta eteenpäin.

Kahden vuosikymmenen mittainen tutkimusmatkani olisi jäänyt kesken ilman joitakin avainhenkilöitä. Vielä 2000-luvun alussa eivät käytettävissä olevat konkordanssityökalut tuottaneet aineistostani tarvitsemiani tuloksia. En voi kyllin kiittää silloisen kieliteknologian opiskelijan Miikka Silfverbergin apua tarvitessani kaikenlaisia skriptejä aineiston käsittelyyn. Tuhannet kiitokset, Miikka. Jonas Westerlund ohjelmoi konkordansserin uudenlaisia tarpeitani varten aivan työn viimeisessä vaiheessa 2010-luvun lopulla. Kiitos, kultainen Jonas. Kiitän rakasta sisartani Leenaa ja velipuoliani Makea ja Eikkaa sekä kälyäni Stinaa ja kaikkia ystäviäni ja muistan kiitollisuudella edesmenneitä äitiäni Inkeriä ja isääni Niiloa ja appivanhempiani Karin ja Henry Westerlundia ja heiltä saamaani tukea vuosien mittaan. Perheeni on ollut ratkaisevan tärkeässä roolissa tukenani tällä matkalla. Rakkaat puolisoni Anders, tyttärenne Katarina, poikamme Jonas, miniämmme Sandra ja ennen kaikkea mummon silmäteriä Leon ovat pitäneet huolta ilosta ja onnesta matkan varrella ja kollokaatiopeikot kaukana. Kiitos, ihanat.

Helsingissä elokuun 20. päivänä 2020, Hanna Westerlund

Sisältö

1 Johdanto	1
1.1 Tutkimuksen tavoite, tutkimuskohde, hypoteesit ja tutkimuskysymykset	1
1.2 Kollokaation teoreettinen viitekehys	5
1.3 Deskriptiivinen tutkimusote	8
1.4 Heräte tutkimusaineiston koostamiseen	10
1.5 Tutkimuksen kulku, painopistealueet ja rakenne	11
2 Kollokaatio kontekstissaan	15
2.1 Kollokaation käsite	20
2.1.1 Kollokaation rakennetta ja merkitystä kuvaavien osien nimet	21
2.1.2 Kolligaatio ja myötäesiintymä	22
2.1.3 Yhteisesiintymä ja vakiintuma	23
2.2 Näkökulmia kollokaatioon ja sen ominaispiirteitä	23
2.2.1 Toistuvuus, tuttuus, tyypillisuus, todennäköisyys	27
2.2.2 Rakenne	31
2.2.3 Merkitys/intensio	37
2.2.4 Mielleyhtymä/assosiaatio	47
2.3 Konventionaalisuus ja norminmukaisuus	54
2.4 Tekstiyhteys ja tilannekonteksti	55
2.5 Merkityksen kontekstisidonnaisuudesta	57
2.6 Vapaan ja idiomaattisen valinnan periaatteet	58

2.7	Yhteenvedo: kollokaatio	60
3	Erikoiskieli, säädöskieli, suomennos	63
3.1	Oikeuskieli erikoiskielenä	63
3.2	Virkakieli, säädöskieli ja lakikieli	65
3.3	Terminologisia näkökulmia	66
3.4	Käännöstieteellisiä näkökulmia	68
3.5	Autenttinen erikoiskielen aineisto	72
3.6	Yhteisesiintymien rekisterisidonnaisuus	76
3.7	Käännökset tutkimuskohteena	76
3.8	Säädöskielen ja EU-kielen tutkimusta Suomessa	78
4	Tutkimusaineistot ja tutkimuskohde	81
4.1	Tutkimusaineisto EU ja ensimmäinen verrannaisaineisto SL	82
4.2	Toinen verrannaisaineisto: semfinlex-lait	83
4.3	Kolmas verrannaisaineisto: LaDi	85
4.4	Neljäs verrannaisaineisto: lehdet-osa	87
4.5	Aineistojen koko, edustavuus ja populaatio	87
4.6	Yhteenvedo: tutkimusaineistot	90
4.7	Tutkittava termi: <i>toimenpide</i>	91
4.8	Tutkittavat yhteisesiintyjät: <i>ryhtyä, soveltaa, suorittaa, säätää, tehdä</i> ja <i>toteuttaa</i>	92
5	Tutkimusmenetelmät	95
5.1	Korpuslingvistiikasta kielentutkimuksen menetelmänä	96
5.2	Korpusten koostaminen, käsittely ja arkistointi	97
5.3	Tutkimusmenetelmän valinta	100
5.4	Tutkimusmenetelmä prosessina	101
5.5	Tilastolliset menetelmät	103
5.6	Tekstikorpuksen käsittelyn työkalut	114

Sisältö	11
5.7 Yhdyssanojen kartoituksen menetelmä	118
5.8 Tutkimusikkunan laajuus	118
5.9 Epäkelvot hakutulokset	119
5.10 Yhteenvedo ja tutkimusosion rakenne	120
6 Määrälliset ja rakenteeseen liittyvät havainnot: <i>toimenpide</i>	123
6.1 Määrälliset havainnot (EU ja SL)	124
6.1.1 Tutkimuskohde <i>toimenpide</i>	125
6.1.2 Tutkimusverbit yhteisesiintyjänä	127
6.1.3 Yhdyssana yhteisesiintyjänä	142
6.1.4 Yhteisesiintyjien jakaumat	145
6.2 Muutos säädösteksteissä (semfinlex-lait)	147
6.3 Tekstilajikontaisia havaintoja (LaDi)	161
6.4 Havaintoja yleiskielisestä verrannaisaineistosta (lehdet-osa)	168
6.5 Rakenteeseen liittyvät havainnot	172
6.5.1 Yhdyssana yhteisesiintymänä	172
6.5.2 Läheisyys, vierekäisyys, järjestys	175
6.5.3 Kahden yhteisesiintyjän rakenne	177
6.5.4 Verbikantaiset substantiivit	178
6.5.5 Kieliopillinen luku	181
6.5.6 Epäsuorat objektit	183
7 Merkitykseen ja assosiaatioon liittyvät havainnot: <i>toimenpide</i>	185
7.1 Käännösvastineita EU:n IATE-termipankissa	187
7.2 Yhdyssana yhteisesiintymänä	190
7.3 Semanttinen preferenssi ja semanttinen prosodia	194
7.4 Assosiaatio ja vastavuoroisuus	195
8 JOHTOPÄÄTÖKSIÄ JA NÄKÖKULMIA	201
8.1 Tutkimuskysymykset ja tutkimusmenetelmät	202

8.1.1	Tutkimuskysymyksiin vastaaminen	202
8.1.2	Tutkimusmenetelmien arviointia	209
8.2	Ajatuksia tutkimuksen merkityksestä ja jatkotutkimusaiheista	213
8.3	Kollokaatio kääntäjän työkaluksi	216
8.4	Lopuksi	219
Viitteet		227

Luku 1

Johdanto

1.1 Tutkimuksen tavoite, tutkimuskohde, hypoteesit ja tutkimuskysymykset

Suomi liittyi Euroopan yhteisöjen jäseneksi 1.1.1995. Liittyminen edellytti siihen asti säädettyjen EY:n asetusten ja direktiivien saattamista osaksi Suomen lainsäädäntöä ja massiivista käännöstyötä niiden tuottamiseksi suomen kielelle lyhyen ajan sisällä. Suomessa virisi tuolloin keskustelu tekstien ymmärrettävyydestä ja asiasta käytiinkin vilkasta kirjeenvaihtoa yleisönosastoilla, jossa käsiteltiin jäsenyyden hyviä ja huonoja puolia ja liittymisestä järjestettyä kansanäänestystä yleisemminkin. Euroopan yhteisöjen tekstien ymmärrettävyyttä arvioitiin myös valtionhallinnossa, tuloksena opetusministeriön koulutus- ja kielipolitiikan osaston julkaisusarjassa vuonna 1997 ilmestynyt Pirjo Karvosen teos *Suomi Eurooppalaisessa kieliyhteisössä*. Siinä Karvonen toteaa muun muassa että ”eurotekstien käännökset ovat luomassa kieleemme ja kulttuuriimme uutta traditiota, aikaisemmasta poikkeavaa, myös aikaisempaa traditiota muuttavaa” (Karvonen (1997): 7).

Eurotekstit, toisin sanoen Euroopan yhteisöjen tekstit, tuntuivat monista oudon vierailta, mutta syytä oli vaikea yksilöidä – johtuiko se Suomeen rantautuvista eurooppalaisen kulttuurin ja yhteistoiminnan uusista toimintatavoista ja käsitteistä vai oliko kyseessä kielestä nouseva outouden ja vierauden tunne? Jos kyseessä on edes osittain jälkimmäinen, voisiko sitä selittää käytettävissä olevien termien ja/tai kääntäjien sanavalintojen, kollokaatioiden, kautta? Vuonna 1998 ilmestyneessä sanakirjassa *Concise Oxford Companion to the English Language*¹ Tom McArthur toteaa kollokaatiosta, että vikaan osuvat valinnat ovat tärkein vierauden osoitus:

¹Concise Oxford Companion to the English Language. 1998. Retrieved May 31, 2010 from Encyclopedia.com: <http://www.encyclopedia.com/doc/1O29-COLLOCATION.html>

”Collocation is basic to language. Its subtleties must be learned, and failure to get the collocations of English right is a major indicator of foreignness: for example, talking about *rotten* rather than *rancid butter*.”

Samoihin aikoihin heräsi käännöstieteen parissa kiinnostus käännöskielen erikoispiirteitä kohtaan, jolloin käynnistyi muun muassa tämän tutkimuksen näkökulmasta oleellinen Anna Maurasen johtama käännösuniversaalien tutkimushanke korpusaineistoilla (tarkemmin alaluvussa 1.5) Joensuun yliopiston Savonlinnan kääntäjänkoulutuslaitoksella ja korpustutkimuksen mahdollisuudet käännöstekstien tutkimuksessa tulivat myös suomalaisten käännöstieteilijöiden tietoisuuteen ja sen menettelmät tutkijoiden ulottuville.

Näiden kahden tapahtumasarjan tuloksena syntyi ajatus tutkia suomennetun säädöstekstin sisältämiä kääntäjän sanavalintoja, kollokaatioita, korpustutkimuksen menetelmiä hyödyntäen ja verrata niitä alun perin suomeksi kirjoitettujen säädöstekstien vastaaviin valintoihin. Taustalla oli toive siitä, että sanavalintojen tutkimus voisi osaltaan selittää yksilön kokemusta kielen tuttuudesta tai ehkä paremminkin outoudesta: miksi jokin sanavalinta tuntuu niin erikoiselta, että arvelemme puhujan tai kirjoittajan olevan muun kielen kuin suomen syntyperäinen puhuja? Miksi emme useinkaan pysty perustelemaan tarkemmin, miksi jokin asia vain ”ei tunnu oikealta – ei kuulosta suomalta”?

Varsinainen tutkimusaineisto koostuu Suomen Euroopan yhteisöihin liittymisen aikaan suomennetuista säädösteksteistä. Tutkimuksen ensimmäinen ja pääasiallinen verrannaisaineisto on tutkimusaineistoa sisällöllisesti vastaava otos siihenastisesta Suomen säädöskokoelmasta. Tutkittavaksi kielelliseksi elementiksi valikoitui jo tutkimuksen alkuvaiheessa substantiivi *toimenpide* (tutkimuskohteen valinnasta ks. alaluku 4.7). Tutkimuskohteen tekstiympäristöstä tutkitaan tarkemmin verbejä *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä* ja *toteuttaa* ja tutkimusaineistosta saatuja tuloksia verrataan paitsi ensimmäiseen verrannaisaineistoon, myös Kielipankkiin talletettuun kolmeen verrannaisaineistoon. Ne ovat Finlex-tietokantaan perustuvan Semanttisen Finlexin aineistosta koottu Kielipankin osakorpus eduskunnan alkuperäissäädöksistä ja siitä edelleen tähän tutkimukseen valitut vuosikymmenet, Kotuksen korpukseen pohjautuva säädöskielen aineisto, joka sisältää lakeja, asetuksia ja direktiivejä sekä yleiskieltä edustava aikakaus- ja sanomalehtiä sisältävä aineisto. Kolmesta viimeksi mainituista aineiston tavoitteena on tarkastella varsinaisesta tutkimusaineistosta tehtyjä havaintoja toisessa ajallisessa kontekstissa tai genressä. Aineistot on kuvattu tarkemmin luvussa 4.

Tämän sanavalintoihin keskittyvän tutkimuksen teorettinen perusta on John Rupert Firthin (1890–1960) ajatuksissa kontekstin merkityksestä kielellisten elementtien valinnassa. Firth toimi yleisen kielitieteen professorina Lontoon yliopistossa (University of London) vuosina 1938–1956 ja julkaisi vuonna 1957 kokoelman

luentojaan, puheitaan ja artikkeleitaan otsikolla ”*Papers in Linguistics 1934–1951*” (Firth (1957)). F.R. Palmerin toimittama kokoelma ”*Selected Papers of J.R. Firth (1952–1959)*” ilmestyi vasta Firthin kuoleman jälkeen vuonna 1968 (Palmer (1968)). Firthin tieteellinen työ ei aikanaan saavuttanut kansainvälistä kielentutkimuksen yhteisöä, joka jätti hänet pitkälti omaan arvoonsa (Pandit (1970)). Firthin oppilaat M.A.K. Halliday ja John Sinclair jatkoivat työskentelyä hänen teorioidensa pohjalta kehittäen neo-firthiläisenä suuntauksena tunnettua kielentutkimuksen haaraa eteenpäin (esim. Halliday (1994) ja Sinclair (1991)). Erityisesti Sinclairin työ korpuslingvistiikan nostamisessa käännotieteen tutkimuksen menetelmäksi on tutkimukseni keskiössä. Kieli nähdään perusolemuksestaan kontekstuaalisena ja kielelliset valinnat määräytyvät viestintätilanteessa merkityksen muodostumiseen vaikuttavien tekijöiden motivoimina strategisina valintoina. Kieli on inhimillistä toimintaa ja siten Hallidayn termein ”kattavaa, tuhlailtavaa, horjuvaa, ei-autonomista ja varioivaa” (Luukka (2002): 99). Tutkimuksen tavoitteena on osaltaan tarkastella kysymystä, päteekö Hallidayn ajatus myös tarkasti säänneltyyn ja ohjeistettuun erikoiskieleen, jonka tulisi olla yhdenmukaista ja yksiselitteistä.

Firthin työn merkitys tutkimuksessani realisoituu kollokaation käsitteen määrittelyssä. Tutkimukseni pyrkiikin noudattamaan firthiläisen tutkimustradition suuntaviivoja. Kielimalliin liittyvänä tehtävänä onkin kollokaation käsitteen tarkastelu ja määrittely tutkimuksen empiirisen osan tarpeita varten. Käännotteoriat ja erikoiskielten tutkimus tarjoavat käsitteistön tarkasteltavana olevan aineiston erityispiirteiden tarkasteluun ja korpuslingvistisiä menetelmiä tarvitaan tutkittavien kielellisten yksiköiden valintaan ja niiden etsimiseen aineistosta. Tutkimusaineistojen leksikaalisen aineksen vertailu kollokaatioanalyysin avulla on luonteeltaan deskriptiivistä: havainnot tehdään mahdollisimman edustavan aineiston pohjalta mahdollisimman objektiivisesti.

Firthin tallennetuista puheista tai teksteistä ei juuri löydy tutkimustyössä tarvittavia käytännöllisiä kollokaation määritelmiä, kuten vastauksia keskeisiin kysymyksiin: Mitä usein esiintyminen tarkoittaa käytännössä? Kuinka monesta elementistä kollokaatio voi koostua? Miten kaukana toisistaan hakusana ja sen kollokaatti voivat sijaita? Onko kompositionaalisuus tai oikeammin ei-kompositionaalisuus oleellista yhteisesiintymisen määritelmässä? Näihin kysymyksiin on saatu vastauksia vasta kuluvalle vuosisadalla, eikä niistä vielä ole syntynyt yksimielistä ja selkeää käsitystä. Kielentutkimuksen eri osa-alueella ilmiötä tarkastellaan alan omista lähtökohdista ja tarpeista käsin; tämän tutkimukseni näkökulma on kääntäjän ja sen yhtenä tavoitteena on käsitteen määrittäminen siten, että se paremmin palvelee kääntäjän tarpeita.

Pyrin havainnoimaan säädöskielen kollokaatioita Euroopan yhteisöjen säädösten suomennoksissa ottaen huomioon kääntäjän käännottehtävässään tekemät strate-

giset valinnat ja kääntäjän työn reunaehdot. Tavoitteena on tarkastella kysymystä siitä, onko tutkimuskohde muodostanut kollokaatioita tutkittavien verbien kanssa Suomen lainsäädännön teksteistä jo vanhastaan vai saapuuko kansainvälisen yhteistyön mukana kollokaatioita, jotka viestivät tutkimuskohteen saamista uusista merkityssivallöistä säädöskielessä. Määrittelen seuraavassa työn tutkimushypoteesin ja tutkimuskysymykset.

Tutkimushypoteesi:

Euroopan yhteisöjen lainsäädännön rantautuminen Suomeen on tuonut mukanaan paitsi uusia ilmiöitä, myös uusia tapoja ilmaista vanhoja käsitteitä. EU-asetusten suomi on (leksikaalisten) kollokaatioiden osalta erilaista kuin autenttinen, alun perin suomeksi kirjoitettu säädöskieli.

Tutkimuskysymykset:

1. Onko Euroopan yhteisöjen (nyk. Euroopan unionin) asetusten suomentamisen seurauksena syntynyt epätyypillisiä tai uusia (leksikaalisia) kollokaatioita Suomen säädöskieleen?
2. Mitä mahdollisille epätyypillisille kollokaatioille on tapahtunut ajan mittaan? Onko Suomen säädöskieleessä tapahtunut muutoksia niiden seurauksena?

Näiden kahden tutkimuskysymyksen lisäksi tutkimuksen läpi kulkee ajatuksia tai hypoteeseja oletettujen epätyypillisten kollokaatioiden syistä. Tutkimus rakentuu määrällisten havaintojen varaan, joiden pohjalta tuloksia on arvioitu myös kollokaation rakenteen ja merkityksen kannalta. Koska kääntäjiä ainakin nykyisin ohjaavat institutionaaliset konventiot ja vakiintuneet vastineet ja niiden pohjalta rakentuneet termipankit ja konekäännösohjelmat, eivät tutkimusaineistot ja verrannaisaineistot ole keskenään vertailukelpoisia siinä mielessä, että tuotetuista teksteistä voitaisiin hahmottaa, ovatko käänösprosessin seurauksena mahdollisesti tuottuneet epätyypilliset kollokaatiot käänösprosessiin tai tilannekontekstiin liittyviä. Määrällisten havaintojen varjoon jäävät yksittäiset esiintymät saattavat kuitenkin tarkoittaa kääntäjän valintoja työkalujen tarjoamien ratkaisujen ulkopuolella. Käänös-toimessa tapahtuu sekä kääntäjän tekemien valintojen siirtymistä työkaluihin että toisinpäin. Pyrin sen vuoksi tarkastelemaan myös kääntäjien käyttämän termipankin tarjoamia tutkimuskohteen vastineita ja tarkastelen näin kysymystä siitä, millaisia vaihtoehtoja kääntäjällä on nykyisin käytettävissään ja millaista variaatiota vastineet tarjoavat.

Tutkimuskirjallisuudessa kollokaatio on määritelty eri tavoin riippuen taustalla vaikuttavasta tutkimustraditiosta, tutkimuksen kohteena olevasta kielestä sekä tutkimuksen tavoitteista. Oma käsitykseni kollokaatiosta ilmiönä on rakentunut tutkimuksen kestäessä pala palalta, ja usein kohtaamani kollokaation kuvaus on vaikuttanut olevan oman työni kanssa ristiriidassa. Tämä tutkimus kuvastaa omalta osaltaan käsitteen määrittelyn ja rajanvedon haasteellisuutta ja kontekstuaalisuutta. Tarkastelen kysymystä kollokaatiosta ilmiönä tarkemmin luvussa 2.

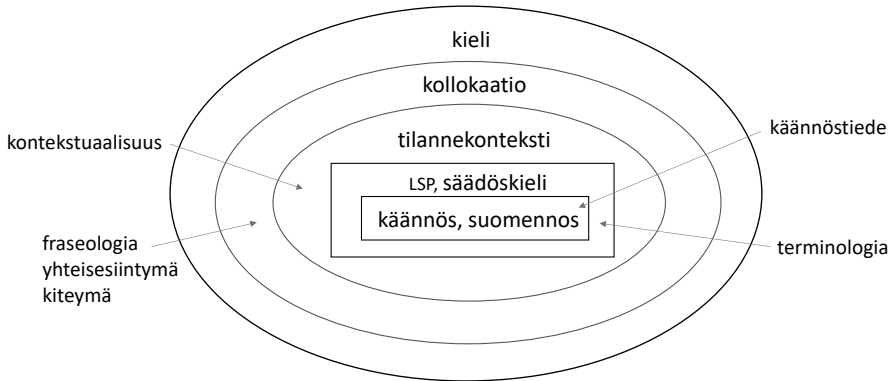
Tutkimuksen perusta on neo-firthiläisessä tutkimustraditiossa. J.R. Firthin artikkelit, puheet ja luennot ovat ohjanneet tutkimustani alusta alkaen niin voimakkaasti, että kun funktionaalisen kieliopin asiantuntija Susanna Shore piti hänen elämäntyöstään esitelmän (Shore (2004)) yleisen kielitieteen laitoksen klassikkolukupiirissä vuonna 2004, huomasin, että kaikki tutkimukseni keskeiset näkökulmat tulivat esitelmän aikana mainituiksi: tutkittavan aineiston autenttisuus, deskriptiivisen tarkastelun painottaminen, erikois- ja ammattikielet tutkimuskohteena, kääntäminen, sanasto ja kollokaatio. Firthin merkitys tutkimuksen kannalta näyttäytyi koko laajuudessaan. Tutkimuksen tavoitteena olikin pitkään pyrkiä seuraamaan Firthin kartoittamia polkuja, kunnes kollokaation käsitteen koko kirjo ja sen moninaisuuden merkitys tutkimuksen toteuttamisen kannalta pakottivat realistisempaan ja konkreettisempaan tavoitteenasetteluun. Firthin työn keskeisin kontribuutio oman tutkimukseni kannalta on vahva näkemys kielen kontekstuaalisuudesta. Palaan tähän tarkemmin seuraavassa alaluvussa 1.2.

Tarkastelen seuraavissa alaluvuissa tutkimuksen teoreettisia lähtökohtia. Jo mainitun alaluvun 1.2 esittelemän teoreettisen viitekehäksen lisäksi esittelen deskriptiivisen tutkimusotteen alaluvussa 1.3, autenttisen kielen merkityksen tutkimusaineistona alaluvussa 1.4 ja lopuksi kuvaan alaluvussa 1.5 tutkimuksen kulun, ne teoreettiset painopistealueet, joita käsittelen työssäni tarkemmin, sekä työn rakenteen.

1.2 Kollokaation teoreettinen viitekehys

Tässä alaluvussa tavoitteeni on asettaa tutkimuskohde teoreettiseen viitekehykseen ja pohjustaa näin näkemyksiäni kollokaation käsitteestä. Ne luovat perustan myös tutkimuskohdetta rajaavien ominaispiirteiden valinnalle. Samalla pyrin motivoimaan valintojani sen suhteen, mitä kielentutkimuksen piiriin lukeutuvista aloista käsittelemän tarkemmin tutkimuksen teoreettisessa osassa.

Kuvassa 1.1 esitän tutkimukseni teoriakentän. Tutkimuksen keskiössä on tekstimuotoinen suomennettu kieli, käännös. Suomennoksia koskeva tutkimus tarjoaa tutkimuksen taustaksi ajatuksia suomennosten ja alun perin suomeksi kirjoitettujen tekstien eroavaisuuksista ja samuudesta (Jantunen (2004), Mauranen (2000)) sekä



Kuva 1.1: Tutkimuksen teoreettinen viitekehys.

ideologian ilmentymänä (Kemppanen (2008)). Tutkimusten anti omalle työlleni on tien avaaminen käännössuomen tutkimuksen menetelmiin ja tutkimustulosten viitoittama tie käsitteen muodostuksessa. Tutkimuksen lähtöoletukset kontekstin merkityksestä kielen rakentumisessa ovat saaneet vahvistuksia näiden tutkimusten myötä.

Tutkimukseni tekstilaji on säädöskieli. Erikoiskielten tutkimukseen liittyy vahvasti terminologian näkökulma, jota tutkimuksessa tarkastellaan tutkimuskohteen termiluonteen vuoksi. Terminologian lisäksi tutkimuksessa tarvitaan fraseologisen tutkimuksen kontribuutiota kollokaation ilmiön kuvaamisessa. Erikois- ja ammattikielissä kollokaatioiden tutkimus liittyykin pitkälti fraseologiaan, katso esimerkiksi Gledhill (1996), Gledhill (2000), Pearson (1998), Heid (2001), Bowker ja Pearson (2002). Tilannekonteksti tutkimuksessani on Euroopan unionin käännöstoiminta, mikä asettaa omat reunaehdonsa suomennosten tutkimukselle.

Käännöstieteen tavoitteista lähellä omaa tutkimustani ovat kääntämisen prosessin kuvaus, kääntäjän työkalujen kehittäminen, lähdetekstin kääntäjän kannalta keskeisten kielellisten elementtien tunnistaminen ja tavoitteena olevaan käännöksen idiomaattisuuteen liittyvät kysymykset, esimerkiksi Lehr (1993), Kenny (2001), Tognini-Bonelli (2001). Kielitieteellinen tai käännöstieteellinen tutkimus tulee toisinaan lähelle omaa tutkimustani olematta kuitenkaan selkeästi sille rinnakkaista kollokaatiotutkimusta. Esimerkiksi psykologivastainen tutkimus, jossa tavoitteena on ha-

vainnoida kielen synnyn mekanismeja yksllötasolla (Hoey (2005)) auttaa selittämään käänöksissä havaittujen epätyypillisyyksien syitä ja synonymian tutkimus liittyy kiinteästi kollokaattien variaation pohtimiseen (Arppe (2008), Jantunen (2004), Vanhatalo (2003)). Käsittelen liittymäkohtia oman tutkimukseeni tarkemmin kollokaation ilmiötä pohtivassa luvussa 2.

Fraseologian ja kollokaatiotutkimuksen välistä yhteyttä olen päässyt pohtimaan aivan tutkimuksen loppumetreillä. Ajattelin pitkään, että fraseologian kenttä on yhtä moniulotteinen kuin kollokaationkin, mutta ettei näillä tutkimusaloilla olisi erityistä yhteistä tarttumapintaa. Kuitenkin fraseologian puolella on otettu kantaa kollokaatioihin ja ne on muun muassa sisällytetty fraseemien luokittelussa yhdeksi erityisluokaksi (ks. Korhonen (2018)). Fraseologian suhdetta kollokaatioon tai päinvastoin käsittelen tarkemmin alaluvussa 2.2.

Kollokaatioita on tutkittu myös kielenoppimisen puolella, katso esimerkiksi Charniak (1997), Lewis (2000). Suomessa on tutkittu korpuslingvistiksi menetelmin myös oppijankieltä fraseologisessa viitekehyksessä, katso tarkemmin Jantunen (2009) ja Jantunen ja Brunni (2012). Informaatiotutkimus puolestaan tarvitsee keinoja esimerkiksi termintunnistuksen tai erilaisten tekstien automaattiseen luokitteluun tai hakukoneiden toiminnan tehostamiseen, esimerkiksi Järvelin (1995), Järvelin ja Sormunen (2010).

Tietotekniikan kehitys on avannut kollokaatioiden tutkimukselle aivan uudenlaisia näkymiä laajojen kieliaineistojen, korpuslingvististen työkalujen, jäsennessuomenelmien, tilastollista merkitsevyyttä mittaavien menetelmien ja konkordanssityökalujen saatavuuden parantuessa. Korpuslingvistiikan menetelmien vakiintumisen myötä on saatu tutkimustietoa sekä menetelmien soveltuvuudesta kieliaineistojen analyysiin että aineistojen soveltuvuudesta vastaamaan esitettyihin tutkimuskysymyksiin. Yksimieliseen kollokaation määritelmään ja kuvaukseen ei kuitenkaan ole toistaiseksi päästy.

Pyrin määrittelemään kollokaation ja sen keskeiset ominaispiirteet mahdollisimman kattavasti. Tutkimuksen keskiössä on ensimmäiseen tutkimuskysymykseen liittyvä määrällinen tarkastelu ja sen reunaehdot (tarkemmin alaluvussa 2.1), jolloin muiden kuin määrällisten kriteerien tarkastelu jää tässä työssä toissijaiseksi. Tavoitteena on löytää merkitykseen liittyviä kollokaation käsitettä selventäviä havaintoja tutkimusaineistoista aineistolähtöisesti. Tarkastelen seuraavaksi tarkemmin deskriptiivistä aineistolähtöistä tutkimusotetta.

1.3 Deskriptiivinen tutkimusote

Tutkimukseni on luonteeltaan deskriptiivinen. Sen tarkoituksena on selvittää, ovatko havainnot tutkimusaineistosta ja sen verrannaisaineistoista määrällisesti keskenään samanlaisia vai erilaisia. Tutkimuskysymyksen indikoimaa merkityksen muuttumista arvioidaan kollokaattien valikoitumisen perusteella, ei niinkään tutkittavien verbikollokaattien (tai kollokaattikandidaattien) laadulliseen semanttiseen analyysiin pohjautuen. Merkityksen muodostumisen arviointi on siten verrattain ohutta ja tutkimuksen tarkoituksena onkin arvioida jatkotutkimuksen tarvetta näiltä osin.

Deskriptiivisen tutkimuksen kohteena on kieli sellaisena, kuin sitä käytetään. Firthin mukaan deskriptiivinen tutkimusote on lähtökohta kielten ja kulttuurien välisten erojen tutkimuksessa (Firth (1957): 173, Palmer (1968): 202); syntaksin, sanaston, fonologian ja fonetiikan osalta katso (Palmer (1968): 5, 175 ja Shore (2004)). Kieltä tarkastellaan tilannekontekstissaan olennaisina kielellisten ja konkreettisten elementtien välisinä suhteina ”as situational relations in a context of situation” (Palmer (1968): 19). Firthin ajattelutavan mukaisesti kielentutkimuksen tavoitteena ei ole etsiä kaikille kielille yhteisiä universaaleja, eikä kielitieteilijän tehtävänä ole yksittäisten sanojen kuvaus irrallaan niiden esiintymisympäristöstä (Chapman ja Routledge (2005): 81). Yleisenä tavoitteena sen sijaan on rajoitettuihin kielen kuvauksiin soveltuva kielimalli:

”The synopsis presents in outline a general linguistic theory applicable to particular linguistic descriptions, not a theory of universals for general linguistic description.’ (Palmer (1968): 202)

Jokainen eri tilanteessa syntyvä lausuma tai ilmaus on erilainen, joten on väistämätöntä, ettei sen analyysi ole yleistettävissä, eikä tilanne välttämättä toistettavissa. Firthin teorian mukaan rakenteellisia analogioita tai universaaleja ei siis ole edes tarkoituksenmukaista etsiä. (Honeybone (2005)). Oletan siis, että erikoiskielissä kielellisiin ratkaisuihin vaikuttavat tilanteessa vallitsevat piirteet ovat yleiskieltä rajoitetumpia. Esimerkiksi säädöskielen prosesseissa viestinnän osapuolet ovat erikois- ja ammattikielten ammattilaisia, kääntäjiä, terminologeja ja muita tekstien tuottamiseen osallistuvia asiantuntijoita sekä tekstejä tulkitsevia ja soveltavia alan asiantuntijoita, esimerkiksi tuomareita ja lainsäätäjiä. Toiminta on itsessään erityisen säädelyä erikoisalan sisäisten toimintamallien vuoksi, joita ovat muun muassa lakien säätämiseen liittyvät konventiot. Erikoiskielistä tehtyjä havaintoja voidaan helpommin yleistää kyseisen erikoisalan muihin kielenkäyttötilanteisiin.

Deskriptiiviselle käännöstutkimukselle on ominaista, että se tarkastelee tekstiä ilman arvottavaa näkökulmaa (ks. esim. Toury (1995)). Käännöstekstit toteuttavat

viestintätehtävää, joka ei ole sen vähäarvoisempi tai arvokkaampi kuin muukaan kielellinen viestintä, mutta ne eroavat alkukielellä kirjoitetuista teksteistä jopa siinä määrin, että käsitteet 'translationese' ja 'käännöskieli' ovat jo osa tieteenalan vakiintunutta käsitteistöä. Kiinnostavia deskriptiivisen käännöstutkimuksen kysymyksiä ovat esimerkiksi noudattaako käännös kohdekielen normeja (Toury (1995)) ja lukijan odotuksia, tuoko käännös mukanaan kohdekulttuuriin uutta kielellistä ainesta (käännöslainoista katso Chesterman (1997)), tai onko se esimerkiksi ideologisesti värittynyttä (ks. esim. Probirskaia (2009), Kemppanen (2008)). Tässä tutkimuksessa eroja kartoitetaan tutkimalla käännetyn tekstin ja alkukielellä kirjoitetun tekstin kollokaatioita ja tehdään niitä koskevia aineistosta nousevia havaintoja.

Käännetyistä teksteistä voidaan nähdä kääntäjän tyypillisiä tai epätyypillisiä strategisia ratkaisuja, jos tyypillisyyden kriteerinä pidetään tämän tutkimuksen tapaan kohteen määrällistä vertaamista muihin vastaavassa tilanteissa tehtyihin ratkaisuihin. Ulkoisena kriteerinä voi olla toisen kääntäjän ratkaisu, esimerkiksi kohdekielen käännösvastineen valinta vastaavassa tilanteessa, alun perin kohdekielellä kirjoitettu teksti tai alan asiantuntijan näkemys. Käännöksen ei myöskään aina tarvitse olla kohdekielen konventioiden mukainen, toisin sanoen käännös, jota ei huomaa käännökseksi. Toisinaan on etuakin siitä, että teksti vaikuttaa käännökseltä. Toury (1995) erottaa toisistaan lähdekielen ja lähdekulttuurin ja toisaalta kohdekielen ja kohdekulttuurin normeja noudattavat käännösstrategiat ja näitä vastaavat adekvaattisuuden (lähdekielen normit oleellisia) ja hyväksyttävyyden (kohdekielen normit oleellisia) käsitteet (Toury (1995): 79). Jos kääntäjä päätyy käyttämän käännöksessä epätyypillisiä sanavalintoja, ne saattavat olla seurausta adekvaattisesta käännösstrategiasta kääntäjän tietoisena valintana.

Ei riitä, että kielioppi määrittelee ne rakenteet, joita lauseeseen voi sisältyä. Kielopin tehtävänä on kuvata rakenteita, joita ihmiset tosiasiaassa käyttävät. Sen tulee tarjota tietoa eri vaihtoehtoisten rakenteiden todennäköisyyksistä, jolloin kielenpuhujalla saa varmuuden siitä, mikä niistä olisi kieliyhteisössä kyseisessä kontekstissa ensisijainen. (Abney (1995): 13). Samaa lähtökohtaa sovellan tutkimukseni kohteena oleviin leksikaalisiin valintoihin. Kääntäjän käyttämät apuvälineet, käännösmuistit ja konekäännös tekevät valintoja myös kääntäjän puolesta silloin, kun niihin sisällytetyt vastinevaihtoehdot sisältävät käännettävään tekstiin sisältyviä kollokaatioita. On mahdollista, että ne tulevat helposti jo valmiiksi hyväksytyinä vaihtoehtoina automaattisesti valituiksi myös silloin, kun sanavalintoihin olisi hyvä kiinnittää erityistä huomiota. Tässä tutkimuksessa tutkimusaineiston aikaisiin kääntäjän apuvälineisiin ei enää ole pääsyä, mutta tämänhetkisiä vaihtoehtoja pyritään tarkastelemaan takaisinkäännöksen ajatusta hyödyntäen (tarkemmin luvussa 7).

1.4 Heräte tutkimusaineiston koostamiseen

Tutkimusaineiston valintaan vaikutti Suomessa 1990-luvun alussa herännyt keskustelu Euroopan yhteisöjen tekstien ymmärrettävyydestä ja tekstejä kohdannut kritiikki, muun muassa Karvonon (1997).

Käännösten laatu riippuu suurelta osalta lähdetekstien laadusta ja viranomaiset Suomessa ovatkin ajan mittaan ottaneet kantaa EU-tekstien selkeyteen. Suomen kielen lautakunta esittää kannanotossaan 1.12.2005, että Suomi nostaisi esiin Euroopan unionin alkutekstien laadun EU:n virkakielen selkeyttämisestä vuonna 1998 tehdyn sopimuksen hengessä. Käännösten laatu paranee alkutekstien laadun paranemisen myötä. Lautakunta huomauttaa myös, että Euroopan komissio on kiinnittänyt asiaan huomiota ja antanut asiasta tiedonannon vuonna 2005². Tässä tutkimuksessa yksittäisen asetuksen lähtökieli ei ole tiedossa, koska suomennoksia on työn alkuaikoina todennäköisesti tehty useista Euroopan yhteisöjen suurista kielistä, ranskasta, englannista ja saksasta. Tavallisin lähtökieli oli ranska, mutta myös muista kielistä käännettiin alkuaikoina paljon (Simo Määttä, henkilökohtainen tiedonanto, Lammin TRAST-seminaari 16.8.2016).

Ylen toimittaja kommentoi EU-kieltä radio-ohjelmassa Akilleen kantapää³ vuonna 2012 seuraavasti:

”Sytä kunnallisvaalien alhaiseen äänestysprosenttiin on esitetty monia. Yhden yksinkertaisimmista ja monimutkaisimmista esitti arkkipiispa Kari Mäkinen: poliittinen eliitti on eristänyt itsensä puhumalla kieltä, jota äänestäjät eivät enää ymmärrä.

Mäkisen mukaan ”Nyky päivänä monille jäävät hämäräksi vaikkapa talouden EVM ja ERVV sekä hallintokielen Sote eli asiat, joihin yhteisiä rahojamme pannaan”.

Ongelma ei ole uusi. Poliitikot elävät ja puhuvat niin paljon keskenään, että heille muodostuu oma slanginsa, joka ei aina aukene muille. Kun valtakunnan politiikan lisäksi on vielä EU:n laajuinen poliittinen kenttä, hallinnollinen jargon sen kun pahenee. Ei ihme, että kansalaisyhteiskunnan jäsenellä saattavat mennä sekaisin EVM, ERVV ja EVVK.”

Tekstien saatavuus vaikutti ratkaisevalla tavalla aineiston valintaan. Korpuksia ei ollut 2000-luvun vaihteessa vielä saatavilla, mutta Suomen liittyessä Euroopan yhteisöihin julkaistiin yhteisöjen säädökset internetissä Eur-Lex-verkkosivustolla EU:n

²https://www.kotus.fi/ohjeet/suomen_kielen_lautakunnan_suosituksia/kannanotot/suomen_vaadittava_selkeita_alkuteksteja_eu-puheenjohtajakaudella

³Akilleen kantapää 7.11.2012 klo 17.20 yle.fi http://yle.fi/aihe/artikkeli/2014/09/11/aristoteleen_kantapaa-ja-evm-ervv-ja-evvk

avoimen tiedotuksen politiikan mukaisesti. Aineisto koottiin internetissä julkisesti saatavissa olevasta kieliaineuksesta, joka muodostui Suomen säädöskokoelmaan tulevaisuudessa sisällytettävistä säädöksistä siinä vaiheessa, kun niitä ei vielä ollut liikaa kopioitavaksi. Korpus sisältää sen koostamisajankohtana lähes kaiken saatavissa olevan asetuksista koostuvan kieliaineksen.

1.5 Tutkimuksen kulku, painopistealueet ja rakenne

Tutkimukseni käynnistyi osana professori Anna Maurasen johtamaa Joensuun yliopiston hanketta ”Käännösuomi ja kääntämisen universaalit – tutkimus korpusaineistolla”, jossa tavoitteena oli kartoittaa käännösuniversaaleja käännytyistä ja alun perin suomeksi kirjoitetuista aineistoista. Hankkeen tuloksena ilmestyi kolme väitöskirjaa Savonlinnan kansainvälisen viestinnän laitoksella koottujen käännösuomen aineistoista: Sari Eskolan tutkimus syntetisoivista rakenteista ilmestyi vuonna 2002 (Eskola (2002)), Jarmo H. Jantusen synonymiaa käsittelevä tutkimus vuonna 2004 (Jantunen (2004)) ja Hannu Kemppasen tutkimus avainsanoista ja ideologiasta vuonna 2008 (Kemppanen (2008)).

Tutkimus sijoittuu **käännöstieteen** alaan. Kokemukseni mukaan käännösteorioiden tietoinen soveltaminen käännöstyössä on usein vaikeaa. Sama havainto löytyy Euroopan komission käännösyksikön silloisen esimiehen Emma Wagnerin ja käännöstieteen professori Andrew Chestermanin vuoropuhelusta vuodelta 2002 (Chesterman ja Wagner (2002)). Kääntämisen tutkijat tarkastelevat samankaltaisia kysymyksiä eri näkökulmista, jolloin karttuvalle tutkimukselle tyypillinen käsitteiden moninaisuus ja termien paljous pitävät teoriat vaikeasti lähestyttävänä. Olen omassa työssäni havainnut, että käsitteiden moninaisuus ja niiden käytön päällekkäisyys vaikeuttavat teorioiden tietoista hyödyntämistä. Kääntäjällä ei aina ole mahdollisuutta vaikuttaa työn strategiaan valintoihin tai sitä kautta sanavalintoihin, vaan useimmiten on noudatettava alalla jo vakiintuneita käytäntöjä. Kääntäjä tekee työssään kuitenkin koko ajan tietoisia ja tiedostamattomia valintoja ja kääntämisen lainalaisuuksien tiedostamisen avulla kääntäjä voi nostaa alitajuisia valintojaan tietoiselle tasolle. Kollokaatio voisi olla sellainen kääntäjän tiedostamaton mekanismi, jonka tunnistamisesta saattaisi olla kääntäjälle hyötyä työssään. Tämä tutkimus pyrkii osaltaan selkeyttämään kollokaatioon liittyviä käsitteitä ja kollokaation hyödyllisyyttä kääntäjän työkaluna (Westerlund (2009)). Tarkastelen kääntämisen tuloksena syntyneitä tekstejä, joita ei voi linkittää yksittäisen kääntäjään. Suomen Euroopan unioniin (silloinen Euroopan yhteisö) liittymisen seurauksena tehdystä käännöstyöstä, sen jakautumisesta eri kääntäjien kesken tai käytetyistä apuvälineistä ei ole ollut tämän tutkimuksen ulottuvilla dokumentaatiota, jonka perusteella tilannekontekstia olisi mahdollista näiltä osin kartoittaa. Lähden kuitenkin siitä, että tekstit on asian-

mukaisesti tarkastettu ja hyväksytty suoraan voimaan tulevina asetuksina osaksi Suomen lainsäädäntöä, ja ne ovat sen vuoksi edustavia esimerkkejä tutkimuksen kohteena olevan termin käsittelystä.

Toinen tähän tutkimukseen läheisesti liittyvä ala on **erikoiskielten tutkimus**. Ghis Gledhill perustaa omat näkemyksensä Hallidayn (esim. Halliday (1994)) funktionaaliseen teoriaan maailman rakentumisesta ja tulkinnasta diskurssin kautta. Kielen rakenteet ovat keskeisiä diskurssin rakennusaineita. Kielenkäyttäjät tarvitsee kommunikaatiotilanteessa leksikaalista tietoa konventioista, muun muassa kyseisen erikoisalan konventioiden tai normien mukaisista sanavalinnoista tunnistaakseen niistä poikkeavat valinnat. (Gledhill (1996), Gledhill (2000)). Gledhill käyttää kollokaatioista tämän tutkimuksen käsitteistöstä poikkeavasti termiä fraseologinen yksikkö ja hänen mukaansa tällaiset retoristen tavoitteiden motivoimat, diskurssia rakentavat fraseologiset yksiköt⁴ on opittu tai ne olisi opittava jo aikaisemmin käydyin diskurssin sisällä ja oppimisen tulisi tapahtua yhteisön sisällä. Kollokaatio on eräs tällainen fraseologinen yksikkö. Poiketen Gledhillin ja Tieteen termipankin käsitteistöstä. Tutkimuksessani termejä **fraasi** ja **fraseologia** käytetään Stefan Evertin tekemän jaon (ks. tarkemmin luku 2) mukaisesti merkityksen perusteella rakentuvan kollokaation kuvauksessa. (Gledhill (2000)). Kollokaatiosta fraseologian viitekehkeyksessä tarkemmin alaluvussa 2.2.

Kahdesta erilaisesta käännetyn erikoiskielten aineistosta kollokaatioita tutkinut Bernardini arvelee, että lähtökielen vaikutus teknisten kollokaatioiden (technical collocations) kääntämisessä on suurempi kuin runsaammin kielikuvia sisältävää erikoiskieltä käännettäessä, joka kääntyy useimmin standardisoiduiksi, jopa stereotyyppisiksi kollokaatioiksi (Bernardini (2015): 535). Voidaan olettaa, että kollokaatio ilmiönä vaikuttaa myös erikoiskielissä.

Tutkimukseni taustalla vaikuttavaa kontekstuaalista kielikäsitystä ja sen mukaisista ajatusta merkityksen muodostumisesta kontekstin kautta käsittelen tarkemmin luvussa 2, jossa kuvataan myös työn kohteena oleva ilmiö kontekstissaan sen ominaispiirteiden kartoituksen avulla. Luvun 3 kuvaus säädöskielestä erikoiskielenä ja erityisesti suomennetusta säädöskielestä tutkimuskohteena ja suomalaisen säädöskielen tutkimuksesta taustoittaa tutkimusaineistojen erityispiirteitä. Luvussa 3 kartoitan myös käännostapahtumaan liittyviä käännostieteellisiä näkökulmia. Luvussa

⁴Fraseologinen yksikkö (phraseological unit): Määritelmä: kahden tai useamman sanan muodostama kiinteä paradigmaattis-syntagmaattinen kokonaisuus, Selite: Sanat tai sananmuodot esiintyvät tyypillisesti toisten sanojen, sananmuotojen, kieliopillisten tai semanttisten kategorioiden kanssa kiinteässä vakiintumassa. tällainen sanojen muodostama suhde voi olla esimerkiksi kollokaatio. Tällaisia voivat olla esimerkiksi kielenulkoisesti motivoituneet lehmä ammuu tai hevonen hirnuu, mutta myös monet vakiintumat, joiden motivaatio ei ole kielen ulkoinen, kuten kaunis ilma (mutta ei tavallisesti komea ilma). Tieteen termipankki 11.04.2016: Kielitiede:fraseologinen yksikkö. (Tarkka osoite: http://www.tieteentermipankki.fi/wiki/Kielitiede:fraseologinen_yksikkö.)

4 kuvaan tutkimusaineistot ja tutkittavan lekseemin ja sen kollokaatiokandidaattien valintaprosessin. Kollokaatioiden ominaispiirteiden kartoituksessa olen lisäksi tarvinnut joitakin korpuslingvistiikan menetelmiä, jotka tulevat esitellyiksi luvussa 5. Ensimmäinen analyysiluku 6 sisältää tutkimuskohteen painopisteenä olevan määrällisen analyysin ja rakenteeseen liittyvät havainnot. Toisessa analyysiluvussa 7 kuvaan merkitykseen ja assosiaatioon liittyviä havaintoja, jotka nousevat määrällisten havaintojen tarkastelusta. Luvussa 8 esitän tutkimuksen johtopäätökset ja sen herättämät ajatukset. Taustalla kulkee koko ajan kokemus kollokaation merkityksestä kääntäjän työssä ja pyrkimys havainnollistaa määrällisen tutkimuksen näkökulmaa käännöstieteellisessä tutkimuksessa.

Luku 2

Kollokaatio kontekstissaan

Tarkastelen tässä luvussa kollokaation piirteitä tutkimuskirjallisuuden ja joidenkin sanakirjojen valossa kääntäjän ja käännöstieteen tutkimuksen näkökulmaa painottaen. Yleisenä tavoitteenani on konkreettisia tekstiesiintymiä tutkimalla saavuttaa parempi ymmärrys kollokaatiosta ilmiönä ja tarkastella mahdollisuuksia hyödyntää sitä kääntäjän työssä. Pyrin hahmottamaan, millaisena ilmiö näyttäytyy tarkasti säännellyssä ja ohjeistetussa säädöskielessä. Yleiskielessä meille suomalaisille on itsestään selvää, että *emävale* on *iso* ja jollakin asialla voi olla *suuri merkitys* – tai voisiko merkitys sittenkin olla myös *iso*? Tiedämme myös, että yleensä *ryhdymme toimenpiteisiin* – paitsi lääkärit, jotka tapaavat *tehdä* niitä, tai Euroopan yhteisöjen organisaatio, joka voi niitä myös *säätää* (ks. tutkimustulokset luvussa 6 ja johtopäätökset luvussa 8).

Kokoan tässä luvussa yhteen kollokaation määrittelyyn liittyviä käsitteitä (alaluku 2.1) ja sen ominaispiirteitä ja erilaisia ilmiötä valottavia näkökulmia (alaluku 2.2). Tarkastelen konventionaalisuutta ja norminmukaisuutta (alaluku 2.3), kontekstin käsitettä (alaluku 2.4), merkityksen kontekstisidonnaisuutta (alaluku 2.5) sekä taustalla vaikuttavia vapaan ja idiomaattisen valinnan periaatteita (alaluku 2.6). Tarkasteltavien ominaispiirteiden valinta perustuu sanakirjoissa ja alan kirjallisuudessa esiintyviin määritelmiin ja niistä löytyviin avainsanoihin. Kaikki esitellyt kollokaation ominaispiirteet eivät ole tämän tutkimuksen keinoin tarkasteltavissa, joten käsitelen alaluvuissa myös tutkimuksen edellyttämiä käytännön rajoituksia ja valintoja. On huomattava, että tämä tutkimuksen teoriaosuus pohjautuu muutamaa poikkeusta lukuun ottamatta vahvasti englannin kielen tutkimukseen.

Kollokaatio⁵ voidaan määritellä, nimetä ja luokitella monin eri tavoin. Tutki-

⁵Kollokaatio; Määritelmä: usein yhdessä esiintyvien sanojen muodostama kokonaisuus (Lähde: Tieteen termipankki 2.03.2016: Kielitiede:kollokaatio.). Tässä tutkimuksessa käytän kollokaatiosta ilmiönä termiä yhteisesiintyminen ja sen edustumasta tekstissä termiä yhteisesiintymä. Yhteisesiin-

muksen metakieli, käytetyt käsitteet ja termit ovat niin ikään vakiintuneet eri tutkimushaaroissa ja yksittäisten tutkijoidenkin käyttämässä kielessä erilaisiksi. Kollokaation tai sille sukua olevien ilmiöiden kategorisoinnin kannalta välttämättömiä tai oletettuja ominaispiirteitä löytyy useita. Ilmiö on moniulotteinen, samoin kuin siitä käytävä tieteellinen keskustelukin. (Evert (2008), Seretan (2011)). Fred Karlssonin mukaan tieto kollokaatioista, joita hän kutsuu ’taipumukseksi yhteisesiintymiseen tiettyjen muiden sanojen kanssa’ (Karlsson (2008): 188), on yksi leksikaalisen tiedon lajeista. Enemmän tai vähemmän synonyymisiä termejä ovat esimerkiksi yhteisesiintymä, vakiintuma, sanaliitto⁶ ja fraseologian puolella esimerkiksi moniosainen ilmaus (MWE; multi-word expression), moniosainen yksikkö (MWU; multi-word unit) tai leksikaalinen fraasi (lexical phrase).

Sagin ym. kollokaatio (collocation) kattaa mahdollisimman laajan merkityksen, jolloin kollokaatio on ”any statistically significant cooccurrence, including all forms of MWE as described above and compositional phrases which are predictably frequent (because of real world events or other nonlinguistic factors)” (Sag et al. (2002): 8). Taustateoriasta, kielentutkimuksen osa-alueesta ja tutkimuksen tavoitteista ja rajauksista riippuen kollokaatio voi terminä palvella yläkäsitteenä, alakäsitteenä, abstraktina ilmiönä tai sen konkreettisena ilmentymänä, reaalistumana tekstissä tai puheessa, jolloin se on väistämättä määritelty eri tavoin monenlaisista lähtökohdista (Biber (2015), Gries (2015), Herbst (2012), Seretan (2011)). Kollokaation määrittelyyn vaikuttavat siis tutkimuksen sovellusala, näkökulma ja tutkijan taustalla vaikuttava kielitieteellinen näkemys. Myös tutkijan käsitys siitä, mikä edustaa evidenssiä kielellisten ilmiöiden tutkimuksessa, riippuu valitusta taustateoriasta (Schönefeld (2011)).

Käsitteen (kieli)tieteellinen määrittely ei ole **korpuslingvistiikan** piirissä yksinkertaista tai yksimielistä, kuten Stefan Gries totesi vuonna 2013:

“Perhaps the most famous quote in corpus linguistics is Firth’s (1957:179) “You shall know a word by the company it keeps”. Thus, the notion of collocation, or more generally co-occurrence, has now been at the centre of much corpus linguistic work for decades. As is so often the case, however, this does not

tymä on neutraali käsite, joka ei sisällä oletusta ei-kompositionaalisen merkityksen muodostumisesta.

⁶Sanaliitto; Määritelmä: kahden tai useamman sanan ryhmä, jonka merkitysseikat liittävät yhteen, Selite: Esimerkiksi suomen kieli. Suomen kielessä sanaliiton ja yhdyssanan ero on monesti häilyvä; se on selvä vain kirjoituksessa (esim. nuorimies nuori mies). engl. co-occurrence. Lisätiedot: VISK2008 Sanaliitto on kahden tai useamman sanan muodostama rakenne. Termiä käytetään toisaalta sanaliiton ja yhdyssanan välisestä erosta esim. tapauksissa mummon mökki ja mummon-mökki (>> § 402), toisaalta vakiintuneista sanaliitoista, kuten kynsin hampain, niine hyvineen ja saada selvä. Nimitys sanaliitto ei ota kantaa siihen, onko rakenne lauseke vai ei. (Lähde: Tieteen termipankki 2.03.2016: Kielitiede:sanaliitto.)

mean that we as a field have arrived at a fairly unanimous understanding of what collocations are (in general), how they are best retrieved/extracted, how their strength or other characteristics are best measured/quantified, etc.” (Gries (2013): 135) [Firthin sitaatissa alun perin mukana oleva huutomerkki puuttuu lähteestä]

Ennustettavissa olevaa, tilastollista yhteisesiintymistä (statistical co-occurrence) mitataan keskinäisen odotusarvon (mutual expectancy) avulla (Herbst (2012): 211). Esiintymistajuus ennustaa tilastollista merkitsevyyttä.

Vieraan kielen oppimisen kvalitatiivisessa tutkimuksessa lähtökohtana ja tutkimuksen kohteena on vakiintunut, toisin sanoen institutionaalinen ja rajoitettu leksikaalinen yhteisesiintyminen (established/institutional restricted lexical co-occurrence), joka on kuitenkin usein mahdoton tunnistaa kielen säännönmukaisuuden perusteella: aina ei ole keinoa ennustaa, mikä sanayhdistelmä on idiomaattinen ja mikä ei. Vieraan kielen puhuja tarvitsee tilastollista tietoa yhteisesiintymistä, ellei hän ole riittävästi altistunut tilastolliselle informaatiolle puhumalla ja kuuntelemalla kieltä sen luonnollisissa käyttöyhteyksissä. (Herbst (2012): 211). Kielen oppimisen teorit liittyvät tässä tutkimuksessa assosiaation mekanismien tarkasteluun, josta lisää alaluvussa 2.2.4.

Leksikografian tarpeisiin pyritään tunnistamaan probabeemejä (probabemes), ihmisille luonteenomaisia tapoja sanoa asioita. Tätä aihepiiriä käsitellen tarkemmin alaluvussa 2.2.1, joka käsittelee tavanmukaisuuden ja tuttuuden näkökulmia. Probabeemit ovat kieli- ja kulttuurisidonnaisia. Esimerkiksi 180 päivää voidaan ilmaista englanniksi fraaseilla 'half a year' tai 'six months'. Kumpi versio mahtaa olla tavalisempi englanninkielisen puhujan arjessa? Herbstin tutkimuksessa BNC-aineistosta näistä ensimmäistä löytyy 4009 kappaletta ja toista 49 kappaletta. Saksan kielessä vastaavaa eroa frekvensseissä ei ole (Herbst (2012): 211–212). Havainto painottaa kollokaation luonnetta kielikohtaisena ilmiönä. Tilannekontekstin ja tekstikontekstin merkitys ilmiön kuvaamisessa on eräs tutkimukseni perustavanlaatuisista lähtöoletuksista (ks. alaluku 2.5). Sinclairin työllä on ollut urauurtava merkitys autenttisten esimerkkien hyödyntämisessä sanakirjojen laatimisessa alkaen COBUILD⁷-hankkeesta ja ensimmäisen aineistolähtöisen sanakirjan valmistumisesta. Korpuslingvistiikan menetelmin voidaan sanakirjoihin automaattisesti lisätä muun muassa kollokaatioon ja kolligatioon liittyvää tietoa kielen keskeisistä preferensseistä ja piirteistä (Rundell ja Kilgarriff (2011): 278, Hanks (2008), Hanks (2010)), kunhan tieto ilmiöstä karttuu ja sen myötä menetelmät kehittyvät. Tietous kielen idiomaattisuudesta ja odotuksenmukaisuudesta ja kollokaatioiden roolista niiden tuottamisessa sekä herkkyys tunnistaa sanakirjojen tarjoamista vaihtoehdoista vastineeseen sopivin kollokaatti ovat tärkeä osa myös kääntäjän ammattitaitoa.

⁷<https://en.wikipedia.org/wiki/COBUILD>

Ulrich Heid (Heid (1999), Heid (2001)) määrittelee kollokaation **fraseologisen kollokaatiotutkimuksen** näkökulmasta seuraavasti: ”Collocations are a phenomenon of lexical combinatorics: they involve the lexical, syntactic and semantic properties of lexical items and their syntagmatic cooccurrence” (Heid (2001): 788). Terminologina Heid esittää, että kollokaatio koostuu kahdesta lekseemistä, joista toisen tulee olla termi. Yhdistelmät ovat kahden substantiivin tai adjektiivin ja substantiivin muodostamia lausekkeita. Rajanvedosta moniosaisen termien ja näiden kollokaatioiden välillä hän toteaa, että määrittelyn perusteet ovat erilaisia riippuen siitä, katsotaanko asiaa lingvistin vai terminologin näkökulmasta. Kielitieteilijälle moniosainen elementti näyttäytyy yhdistelmänä, jossa pääsanana merkitys määrittää sen kanssa käytettyjen sanojen valintaa, joka on siten rajoitettu. Tätä Heid pitää osoituksena kollokaatiosta. Terminologi taas on kiinnostunut moniosaisen yksikön denotaatiosta, kysymyksestä siitä, viittaako elementti terminologisesti relevanttiin käsitteeseen (Heid (2001): 791). Terminologiset näkökulmat ovat tutkimuskohteen määrittelemisessä oleellisia ja palaan niihin alaluvussa 3.3.

Heid viittaa Martinin (1992) määritelmään konseptuaalisista kollokaatioista, jotka liittyvät semanttisen valenssin käsitteeseen. Niille on Martinin mukaan tyypillistä, että pääsanana semanttisessa kentässä on jo valmiiksi tilaus tietynlaisille määritelmille. Martinin esimerkki on substantiivi *disease* (’sairaus’), jonka merkitykseen sisältyy oletus luokittelupiirteestä [+CAUSE], ja esiintymä *infectious disease* on siten esimerkki konseptuaalisesta kollokaatiosta. Toisaalta nämä konseptuaaliset kollokaatiot näyttävät olevan ennustettavissa puhtaasti pääsanana semantiikan pohjalta (Heid (2001): 795). Tarkastelen kollokaatiota merkityksen rakentumisen näkökulmasta tämän luvun alaluvussa 2.2.3 ja vastavuoroisuutta kollokaation ominaispiirteinä alaluvussa 2.2.4.

Kielentutkimuksen puolella painotetaan ilmiön sellaisia ominaispiirteitä, jotka auttavat ymmärtämään, rajaamaan ja määrittelemään sitä tarkemmin. Tämän tutkimuksen keskeisin teoreettinen lähde Firthin ja Sinclairin ohella on Stefan Evertin kollokaatiotutkimus (Evert ja Krenn (2003), Evert (2005), Evert (2008)). Hänen mukaansa kollokaatio on ”one of the most controversial notions in linguistics” (Evert (2008): 2). Evert jakaa kollokaatiotutkimuksen kahteen erilaiseen lähestymistapaan, distributionaaliseen ja intensionaaliseen, kun taas esimerkiksi Herbst (2012) vastaa vastaviitteenä tilastollisesti orientoituneeseen (statistically-oriented) ja merkitykseen sitoutuneeseen (significance-oriented) lähestymistapaan (Herbst (2012): 30). Suurin osa Seretanin kokoamista kollokaation (co-occurrence) määritelmistä edustaa distributionaalista (tilastollista) lähestymistapaa (statistical approach) (Seretan (2011): 10).

Evertin **intensionaalinen** (linguistically motivated) määritelmä lähtee liikkeelle lingvististä lähtökohdista (Seretan (2011): 10–13). Tästä näkökulmasta katsottuna kollokaatio on ”intensionally defined concept that does not depend on corpus

frequency information” (Evert (2008): 17). Intensionaalisen lähestymistavan mukaisesti ei ole oleellista, esiintyykö kielellisten elementtien yhdistelmä kieliaineistossa jonkin tietyn vähimmäisesiintymistajuuden verran tai onko sen esiintyminen todettavissa tilastollisesti merkitseväksi. Sen sijaan keskeistä on selvittää, onko kollokaatio kompositionaalinen, millaista modifikaatiota sen rakenne sallii – voidaanko yhdistelmän elementtejä korvata toisilla, muokata tai erottaa yhdistelmästä ilman, että merkitys muuttuu (Manning ja Schütze (1999): 184), Herbst (2012): 32–33). Intensionaalinen lähestymistapa ja fraseologia ovat käsitteellisesti lähellä toisiaan, eikä niiden välinen rajanveto vaikuta mitenkään selkeältä – Gray ym. toteavatkin, että fraaseja voi pitää laajentuneina kollokaatioina (extended collocations), toisin sanoen vähintään kolmen sanan leksikaalisina ilmauksina (Gray ja Biber (2015)).

Kuten Evert huomauttaa, on lähestymistapojen välillä päällekkäisyyttä. Ne ovat erilaisia, mutta myös monella tapaa samankaltaisia (Evert (2008): 16). Molemmille lähestymistavoille yhteistä on näkemys, että kielellinen tuotos (language text/discourse) ei ole tulosta jonkinlaisesta kielioppisääntöjen päälle lisätyistä sanoista, eikä kieliopillisuus ole ainoa leksikaalisen valinnan rajoite (restraint) (Sinclair (1991): 109). Kääntäjän kannalta tämä tarkoittaa sitä, että ns. sanasanainen käänнос saattaa johtaa syntaktisesti idiomaattiseen, mutta sanavalintojen suhteen epäluontevaan (ja epäluotettavaan?) tai epätavalliseen lopputulokseen (Herbst ja Götz-Votteler (2007): 212). Evert käyttää itse termiä **collocation** viittamaan sekä ilmiön yläkäsitteeseen (kollokaatio) että sen lajiin (intensionaalinen kollokaatio) ja viittaa distributionaaliseen lähestymistapaan termillä **co-occurrence**, joka on siten kollokaation toinen laji, tyyppi tai olomuoto. Evertin käyttämä termistö on omiaan aiheuttamaan terminologista hämmennystä, joten tässä tutkimuksessa asia on ratkaistu antamalla molemmille suunnille omat nimityksensä yhteisesiintymä ja vakiintuma, joista tarkemmin osiossa 2.1.3.

Tämän tutkimuksen näkökulmasta on kiinnostavaa, onko distributionaalinen tieto itsessään kääntäjälle tärkeää, vai onko vasta merkitykseen liittyvä tieto arvokasta. Noudatan Evertin jaottelua, mutta varaan tässä tutkimuksessa termin **kollokaatio** yläkäsitteelle ja käytän sen lajeista termejä **yhteisesiintymä** (co-occurrence) ja **vakiintuma** (intensional collocation). Tutkimuksen painopiste on distributionaalisisessa analyysissä ja intensionaalisia havaintoja tehdään distributionaalisen tarkastelun tulosten perusteella. Seuraavassa alaluvussa kuvaan tarkemmin kollokaatioon liittyvän käsitteistön.

Keskeistä oman tutkimukseni kannalta on kollokaation käsitteen tarkentaminen, jotta se soveltuisi kääntäjän tietoisesti käyttämäksi työkaluksi. Palaan **käännöstieteen** pariin luvussa 3. Käännöstieteen piirissä ilmiö on kiinnostanut tutkijoita muun muassa synonymian ja kielten välisten rakenteellisten ja kulttuuristen erojen kartoittamisen sekä käännösuniversaalien näkökulmasta.

2.1 Kollokaation käsite

Kollokaatioon liittyy teknistä termistöä, jota tarvitaan kuvaamaan sen ominaispiirteitä, esimerkiksi kollokaation muodostavien kielellisten yksiköiden lukumäärää ja järjestystä. Tämän alaluvun tavoitteena on selkiyttää käytetty termistö ja antaa lukijalle mahdollisuus tarkastaa tutkimuksen kuluessa käytettyjen ilmaisujen merkitystä.

Termejä tarvitaan kollokaation yleiskieliseen kuvaukseen, distributionaalisen lähestymistavan edellyttämien kollokaation osien kuvaamiseen ja intensionaalisen tiedon kuvauksen tarpeisiin. Tässä tutkimuksessa käytetyt termit on kuvattu alaluissa 2.1.1–2.1.3. On huomattava, että termit on valittu nimenomaan tämän tutkimuksen tarpeisiin sopiviksi, joten muiden vastaavien tutkimusten termien sisältö voi poiketa tässä esitetystä. Termit eivät ole toisiaan poissulkevia vaan täydentäviä: merkitykseen sitoutuvia kollokaatioita ja niiden osia voidaan yhtä hyvin laskea ja tilastoida ja toisaalta tilastollisessa näkökulmassa distribuutioon sitoutuvissaa määritelmässä tarvitaan yhtä lailla merkityksen kuvauksen termejä.

Leksikaalisen tutkimuksen keskeisin käsite on **sana** (word): ”Leksikon perusyksikkö on sana.” (Karlsson (2008): 17). Firth kuvailee ’sanoja’ likipitäen eläviksi olennoiksi, jotka tuijottavat häntä tekstistä: ”They look at us!” (Palmer (1968): 179) ja hän määrittelee kollokaation käsitteen niiden avulla. Kollokaatiot ovat ”actual words in habitual company” (Palmer (1968): 182); (Shore (2004)). Viitatessani tässä tutkimuksessa mihin tahansa sanojen muodostamaan elementtiin, käytän geneerisiä **termejä** sana tai **sanayhdistelmä**.

Kontekstuaaliselle kielikäsitteelle uskollisena Firth pitää kielentutkimuksen pääasiallisena kohteena sanan sijaan **virkeitä** (sentence) (Palmer (1968): 156) ja kehottaa tutkimaan kollokaatiota virkerajat ylittävänä ja puhujien puheenvuorojen välillä toimivana ilmiönä (Palmer (1968): 107). Yhteisesiintyminen voidaan siis esimerkiksi Hallidayn tapaan nähdä myös virke- tai kappalerajojen yli ulottuvana, koheesiota luovana voimana (Halliday ja Hasan (1985)). Olen varsinaisen tutkimusaineistoni ja ensimmäisen verrannaisaineiston käsittelyssä (ks. luku 5) rajannut aineistojen tutkimusalueeksi lause- ja virkerajat ylittävän 11 saneen laajuisen tutkimusikkunan, kun taas toisen, kolmannen ja neljännen verrannaisaineiston tutkimusalue on kokonainen virke. Pohdin tarkemmin mahdollisuuksia havaita lause- tai virkerajoja ylittäviä yhteisesiintymiä johtopäätösluvussa 8.

Seuraavassa kartoitan työn kannalta keskeisiä kollokaation kuvaamisen termejä.

2.1.1 Kollokaation rakennetta ja merkitystä kuvaavien osien nimet

Kielellinen elementti (linguistic element, linguistic item) tarkoittaa tässä tutkimuksessa mitä tahansa konkreettista, rajoitettua osaa kielestä. Se voi olla merkki, sana, tai sane tai niiden yhdistelmä, vaikkapa kokonainen yhteisesiintymä. Kielellinen elementti voi sisältää toisia kielellisiä elementtejä. Käsite on tässä tutkimuksessa kielellisen yksikön tekninen vastine. Monisanaisista rakenteista Firth käyttää nimitystä **piece** (Firth (1957): 122).

Lainaan fraseologian ja moniosaiten ilmausten tutkimuksen puolelta (Baldwin ja Kim (2010)) termin **komponentti** viittaamaan sanayhdistelmän osiin, jotka yhdessä muodostavat kyseisen ilmauksen. Sen rooli tutkimuksessani on painottaa sitä seikkaa, että kyseinen aines on yksi yhdistelmän osapuolista. Sen rinnalla kulkeva (**kielellinen**) **yksikkö** ((linguistic) unit) viittaa koko sanayhdistelmään ja **elementti** konkreettiseen ortografiseen havaintoon. Käsittelen kielellisen yksikön käsitettä tarkemmin osiossa 2.1.1.

Yhteisesiintymä määritellään teknisesti usein **merkkijonona**, jota kieliaineksesta etsitään ohjelmalla tai komennolla. Ohjelmalle annetaan tällöin **hakusana**. Kutsun hakusanan kanssa yhteisesiintymän muodostavaa sanetta yhteisesiintyjäksi ja vakiintuman hakusanan kanssa muodostavaa sanetta vakiintuman kollokaatiksi, koska valitettavasti vaihtoehto vakiintuja tuottaa tässä kontekstissa väärän konnotaation. Sinclair ja korpuslingvistit käyttävät hakusanan nimitystä **noodi** (node), leksikaalisesta yhteisesiintyjästä nimitystä **kollokaatti** (collocate) ja syntaktisesta yhteisesiintyjästä nimitystä **kolligaatti** (colligate) (esim. (Sinclair (1991), Jantunen (2004)). Hausmanin termit kielen oppimisen puolella ovat **Basis** (base) ja **Kollokator** (collocate) (Hausmann 1984: 401, viittaus teoksessa Herbst (2012): 27).

Hakusana tai sen tekstiyhteydessä esiintyvä kielellinen elementti voi olla myös muu kuin sana, esimerkiksi välimerkki tai symboli. Tässä tutkimuksessa **sane** viittaa ortografiseen merkkijonoon, joka voi sisältää mitä tahansa merkkejä, myös välimerkkejä tai symboleja. Tietokoneavusteisessa kielentutkimuksessa aineistosta voidaan tuottaa kielelliset yksiköt **perusmuotoisena** (**lemma**) tai **pintamuotoisena**. Tekstikorpus on yleensä tokenisoitu ja jäsennetty siten, että sen sisältämistä kielellisistä elementeistä on pintamuodon lisäksi tallennettu vähintään lemma ja useimpien myös muuta annotaatiota, kuten syntaktinen kategoria, esimerkiksi sanaluokka, sijamuoto, suku ja luku ja dependenssijäsennettyssä aineistossa myös sanojen välinen dependenssisuhde (aineiston käsittelystä tarkemmin luvussa 5).

Kielellinen yksikkö on suomennos termistä **linguistic unit/listeme**. Se vastaa tässä tutkimuksessa **merkitysyksikköä** (unit of meaning). Termi **yksikkö** viittaa näissä molemmissa palaseen kieltä, jolla on merkitysisältö ja se voi olla merkki, yksittäinen sana tai useammasta elementistä koostuva rakenne. Termiä tarvitaan

tässä tutkimuksessa viittaamaan merkityksellisyyteen: on jokin syy, miksi jokin palkan kieltä on oleellinen rajausta kääntäjän kannalta. Ajatuksen ydin on, että kielellisen yksikön muodostavat kielelliset elementit kuuluvat yhteen ja muistetaan kokonaisuutena semanttisista syistä. Tämän tutkimuksen lähtöoletus on, että kielellinen yksikkö sallii elementtensä pintamuotojen vaihtelua, jolloin saman kielellisen yksikön eri taipumusmuodot edustavat samaa elementtiä: *ryhdymme toimeen* ja *ryhtyi toimeen* ovat saman kielellisen yksikön edustumia. Tämä oletus on tehty tämän tutkimuksen tarpeisiin, eikä sitä voida pitää yleisenä teoreettisena oletuksena. Pääinvastoin: Sinclair ja Carter (2004) esittävät englannin kieltä tarkastelevien tutkimustensa pohjalta, että eri pintamuodoilla ja lauseenjäsenillä on hyvinkin potentiaalisesti erilaisia rooleja ja merkityssisältöjä. Toisaalta Ulla Jäppinen ajatteli vuonna 1999 ilmestyneessä värien nimitysten käyttöä kartoittavassa väitöstutkimuksessaan (Jäppinen (1999)), että esimerkiksi *sininen järvi*, *sinijärvi*, *järvi siintää* ja *järven sini* ovat saman kollokation edustumia (Jäppinen (1999): 30). Kattavalle korpustutkimukselle tämän kaltainen variaatio asettaa aivan omanlaisiaan vaatimuksia.

Jokin yhteisesiintymän tai vakiintuman osista tai elementeistä voidaan nähdä ensisijaisena esimerkiksi siten, että se vaikuttaa muiden piiriinsä sijoittuvien elementtien valintaan. Tällaista kielellistä elementtiä nimitän syntaktisen kielellisen elementin, vaikkapa nominaalifraasin, **pääsanaksi** tai **ydinsanaksi**. Avainsana-analyyseissä tekstistä etsitään **avainsanat** luokittelun tai esimerkiksi tekstin ideologisten painotusten selvittämiseksi. Avainsana-analyysejä on tässä tutkimuksessa tarvittu tutkimuskohteen valinnassa (tarkemmin alaluvussa 4.7).

2.1.2 Kolligaatio ja myötäesiintymä

Firth kutsui kollokaation syntaktista rinnakkaisilmiötä **kolligaatioksi** (colligation) (Firth (1957), Jantunen (2004)). Näen kolligaation kattavan muun muassa sellaisia syntaktisia ilmiöitä kuten valenssi tai rektio, sanajärjestykseen liittyviä kysymyksiä ja esimerkiksi tempuksen ja moduksen ilmaisemiseen liittyviä rakenteellisia kysymyksiä. Rajanveto ei ole kuitenkaan tässäkin tapauksessa selkeää, mikä heijastaa funktionaalisen kieliteorian mukaan syntaksin ja leksikon päällekkäisyyttä ja limittejä, jota Halliday kuvaa termillä 'lexicogrammar' (Halliday (1978)). Tässä tutkimuksessa eroa ei ole tarpeen tehdä, kun huomio ei ole kiinnittynyt syntaktisista syistä valikoituvaan leksikaalisten komponenttien yhteisesiintymiseen tai vakiintumiseen.

2.1.3 Yhteisesiintymä ja vakiintuma

Tässä tutkimuksessa **kollokaatio** tarkoittaa yläkäsitettä, joka sisältää kaksi erilaisista lähtökohdista määrittyvää lajia: syystä tai toisesta muotoutuneen tavanmukaisen sanayhdistelmän sekä intensionaalisen sanayhdistelmän. Ensiksi mainittua kutsum Tieteen termipankkiin kirjatun mallin mukaan **yhteisesiintymäksi** ja jälkimmäisestä käytän nimitystä **vakiintuma** (Piitulainen, henkilökohtainen keskustelu). Vakiintumista lähellä on myös Tieteen termipankista löytyvä käsite **myötäesiintymä**, jolla tarkoitetaan 'kieliopillisen elementin kanssa syntagmaattisessa konstruktiossa esiintyvää muotoa'⁸. Yhteisesiintyminen viittaa tässä tutkimuksessa prosessiin, jossa kollokaatit valikoituvat kvantitatiivisin perustein ja kriteerein, kun taas vakiintuminen tapahtuu prosessissa, jossa kollokaatiot valikoituvat semanttisin perustein, esimerkiksi saman aihealueen käsitteet kollokoivat keskenään: substantiivin *laki* kanssa kollokoivat esimerkiksi *eduskunta*, *lakikirja* ja *tuomari*. Näiden sanayhdistelmien konkreettiset edustumat tekstissä ovat vastaavasti **yhteisesiintymiä** ja **vakiintumia**.

Tässä alaluvussa olen esittänyt ne termit, joita käytän viitatessani kollokaation eri ilmenemismuotoihin ja koostumukseen. Seuraavaksi esittelen erilaisia näkökulmia kollokaatioon ja pohdin niiden soveltuvuutta oman tutkimukseni tarpeisiin. Jotkin näkökulmista ovat tämän tutkimuksen kannalta oleellisia, kun taas toiset eivät tule työn rajausten vuoksi tutkimuksessa käsitellyiksi, vaikka ne ovat kollokaation käsitteen kannalta keskeisiä.

2.2 Näkökulmia kollokaatioon ja sen ominaispiirteitä

Tutkimuskirjallisuuden perusteella syntyy kuva kollokaatiosta monimuotoisena ilmiönä, jonka ominaisuuksia on kielentutkimuksen osa-alueilla monin eri tavoin painotettu, nimetty ja tutkittu (ks. esim. yhteenvedo teoksessa Seretan (2011)). Käsilä olevan tutkimuksen oleellisin näkökulman valintaan vaikuttava tekijä on Stefan Evertin vuonna 2005 tekemä jako toisaalta neo-firthiläisen koulukunnan kollokaation käsitteeseen, jossa lähtökohtana on määrällinen (distributionaalinen) tarkastelu ja toisaalta intensionaalisen koulukunnan teorianmuodostukseen, jossa kollokaatio kiinnitetään vahvasti merkitykseen liittyviin kriteereihin. Tämän tutkimuksen teoriapohja nojaa neo-firthiläiseen tutkimustraditioon ja kollokaation tutkimuksen kannalta se tarkoittaa keskittymistä ensisijaisesti distributionaaliseen tarkasteluun ja

⁸Myötäesiintymä; Määritelmä: kieliopillisen elementin kanssa syntagmaattisessa konstruktiossa esiintyvä muoto. Selite: Esim. juoda-verbin myötäesiintymiä ovat nestemäisiä aineita tarkoittavat substantiivit. (Z. Harris.), engl. co-occurrence (Lähde: Tieteen termipankki 2.03.2016: Kielitiede:myötäesiintymä.)

sen jälkeen tapahtuvaan merkityksen muodostumisen arviointiin. Tutkimustulosten laajuus ja käsittelytapa tässä tutkimuksessa vähentävät mahdollisuuksia merkitykseen liittyvien ominaispiirteiden kattavaan tarkasteluun ja niihin liittyvät havainnot on kirjattu määrällisten havaintojen perustella ja ohessa. Yksittäisten kollokaatioiden merkityksen tutkimus on kääntäjän kannalta toki tärkeää, mutta tässä työssä ensisijainen kysymys on konventioiden muutos ja sitä kautta mahdollisesti syntyvä merkityksen muutos. Merkitykseen liittyvät ominaispiirteet nousevat John Sinclairin tuotannosta, perinteisen semantiikan tai fraseologian viitekehyksen sijaan. Koska kaikkien merkitykseen liittyvien havaintojen pohjalla on määrällinen, aineistolähtöinen tarkastelu, näen tämän tutkimuksen roolin intensionaalisen teorian puolella yksittäisten havaintojen sarjana joistakin merkityksen tutkimukseen liittyvistä ilmiöistä. Palaan käytännön rajanvetoon myöhemmin tässä luvussa.

Ennen ominaispiirteiden esittelyä tarkastelen kollokaation asemaa fraseologian viitekehyksessä. Perinteinen fraseologian tutkimus, jossa määrittelyn lähtökohtana on merkitys (ja jonka pohjalta myös oman tutkimukseni tutkimusasetelma on alun perin määrätynyt), ei ole juuri painottanut laskennallisia menetelmiä tai määrällisiä havaintoja. Uudempi näkemys ottaa kuitenkin uudenlaisen otteen myös merkityksen kontekstuaalisesta roolista ja fraseologinen yksikkö (phraseological unit) tai fraseemi saa muitakin kuin semantiikkaan liittyviä piirteitä. Kehitykseen on vaikuttanut muun muassa leksikaalisen yhteisesiintymisen (lexical co-occurrence) tutkimuksen tuottama tieto kielessä oleellisista ja tavallisista elementeistä, kun taas perinteisessä fraseologiassa huomio on usein kiinnittynyt harvinaisiin elementteihin (Granger ja Paquot (2008): 30–32). Havainnon määrittäminen fraseologiseksi yksiköksi tehdään distributionaalisessa lähestymistavassa eri kriteerein kuin perinteisessä fraseologiasa.

Kollokaatio on usein eräs fraseologisen yksikön olomuodoista: esimerkiksi Jantunen ja Brunni (Jantunen ja Brunni (2012)) sekä Gries (Gries (2008)) luokittelevat kollokaation fraseologiseksi yksiköksi ja Korhonen (Korhonen (2018)) fraseemiksi. Fraseologisilla yksiköillä on eri tutkijoiden määritelmässä piirteitä, jotka ovat enemmän tai vähemmän oleellisia siihen nähden, luokitellaanko jokin havainto fraseologiseksi yksiköksi vai ei. Esimerkiksi Jantunen määritteli vuonna 2009 (Jantunen (2009)) fraseologisen yksikön muun muassa semanttisen vakiintuneisuuden ja jonkinasteisen ei-kompositionaalisuuden avulla siten, että kyseessä on itsenäinen, semanttisesti yhtenäinen kokonaisuus. Hänen määritelmänsä on lähellä Grangerin ja Paquotin (Granger ja Paquot (2008): 35) raportoimaa listaa tavallisimmista erilaisien luokittelujen sisältämistä ominaispiirteistä, joissa keskeisiksi fraseologisen yksikön piirteiksi on tunnistettu 1) syntaktinen rakenne (internal structure), 2) laajuus (extent) liittyen esimerkiksi kielelliseen yksikköön (esim. lauseke tai virke), 3) ei-kompositionaalisuuden aste (degree of semantic (non)compositionality), 4) syntaktinen joustavuus ja kollokoivuus (degree of syntactic flexibility and collocability ja 5)

diskurssifunktio (discourse function). Molemmissa on mukana semanttinen ulottuvuus: onko fraseologisen yksikön keskeinen kriteeri siis edelleen semanttinen? Myös Gries edellyttää omassa katsauksessaan, että fraseologinen yksikkö (phraseologism) on merkitystä luova: "For something to count as a phraseologism, I will require semantic unity, but not non-compositional semantics." (Gries (2008): 6). Gries huomauttaakin, että joissakin fraseologiaa käsittelevissä tutkimuksissa ei-kompositionaalisuus ei ole ollut kriteerinä. Semanttisella yksiköllä (semantic unit) tulee Griesin mukaan olla sanan tai vastaavan tavoin oma itsenäinen merkityksensä. Tämä on Griesin kuudes ja tärkein kriteeri. Viisi ensimmäistä ovat 1) kielellisten elementtien muoto, joka voi olla mikä tahansa leksikaalinen tai syntaktinen elementti (nature of the elements), 2) elementtien lukumäärä (number of the elements), 3) vaadittu frekvenssi aineistossa (number of times that it must be observed), 4) elementtien välinen etäisyys (permissible distance of the elements) ja 5) leksikaalinen ja syntaktinen variaatio (degree of lexical and syntactic flexibility) (Gries (2008): 4–6). Nämä kriteerit täyttävä fraseologinen yksikkö on Griesin mukaan "the co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance" (Gries (2008): 6). Gries kuitenkin pitää kollokaatiota ja kolligaatiota fraseologiaa yksiköitä lähellä olevana ilmiönä (Gries (2008): 18) ja nostaa esiin korpuslingvististen menetelmin löytyviä fraseologiaa yksikköjä lähellä olevia ja niihin rinnastuvia klustereita (word clusters) ja n-grammeja, joihin kollokaatiokin lukeutuu. Tämä kuvastaa kollokaation asemoinnin vaikeutta osaksi fraseologian viitekehystä. Tämän tutkimuksen tarkoituksena on osaltaan selkeyttää kuvaa yhteisesiintymästä ja sen suhteesta fraseologian määritelmiin.

Granger ja Paquot pitävät leksikaalisia kollokaatioita (lexical collocations) yhtenä fraseologian ditributionaalisen kategorian lajina. Distributionaaliset kategoriat fraseologian puolella tarkoittavat heidän mukaansa klusterianalyysin tuloksena löytyneitä kahden tai useamman sanan yhtenäisiä jonoja (continuous sequences of 2 or more words) kun taas yhteisesiintymäanalyysi tuottaa vastaavasti kahden sanan muodostamia sanayhdistelmiä, jotka ovat (tai voivat olla) epäyhtenäisiä (discontinuous combinations of 2 words). (Granger ja Paquot (2008): 39). Jaottelu ei omassa tutkimuksessani ole käyttökelpoinen, vaikka se osoittaa kollokaatioiden ja distributionaalisen lähtökohdan saavan jalansijaa fraseologian piirissä. Leksikaalisten kollokaatioiden lisäksi myös yhdyssanat luokitellaan referentiaalisiksi fraseemeiksi ja ne toteuttavat kielen referentiaalista funktiota (muut kaksi funktiota tässä jaottelussa ovat tekstuaalinen ja kommunikatiivinen). Olen itse ottanut mahdollisimman laajan lähtökohdan kollokaation ominaispiirteiden kuvauksessa, jolloin voidaan Griesin tavoin pohtia väitettä siitä, että kaikki on lopulta fraseologista – tai tässä tapauksessa kaikki kollokoi kaiken kanssa.

Fraseologisen yksikön kriteerit ovat pitkälti samanlaisia kuin tässä luvussa kuvatut kollokaation ominaispiirteet. Yhteistä rajapintaa löytyy esimerkiksi kaikista rakennetta sivuavista kriteereistä, muun muassa fraseologisen yksikön komponenttien lukumäärästä, niiden etäisyydestä toisiinsa, syntaktisesta joustavuudesta ja paradigmaattisen vaihtelun sallimisesta. Määrälliseen havainnointiin keskittyminen ja aineistolähtöiset ajatukset tutkimuskohteesta ja sen potentiaalisista yhteisesiintymistä tarkoittavat kuitenkin tämän tutkimuksen osalta joustamista fraseologisen yksikön tärkeimmästä kriteeristä, joka on tällaiselta yksiköltä edellytettävä merkityksen erikoistuminen suhteessa yksikön komponenttien merkityksiin irrallisina lekseiminä. Tutkimuksen yhtenä tavoitteena on tarkastella yhteisesiintymistä ja sitä, onko komponenttien esiintymisellä yhdessä merkitystä luovaa tehtävää. Käsitteellisesti en rajoita kollokaation ilmiötä merkitykseen liittyvillä reunaehdoilla vaan pidän lähtökohtaisesti kaikkia havaintoja osoituksena yhteisesiintymisestä. Toinen valinta olisi mahdollinen ja tutkimuksen kannalta ehkä hyödyllisempi, mutta se ei ole tämän tutkimuksen puitteissa enää toteutettavissa.

Esitän tässä alaluvussa ne ominaispiirteet, joita katson tarvitsevani yhteisesiintymän käsitteen muodostamista varten ja kartoitan kuhunkin ominaispiirteeseen liittyvää tutkimusta. Tarkoitin ominaispiirteellä sellaista yhteisesiintymän tai vakiintuman ominaisuutta, jonka avulla sanayhdistelmä voidaan luokitella tai tunnistaa yhteisesiintymäksi tai vakiintumaksi ja keskityn niihin ominaispiirteisiin, jota on mahdollista arvioida tutkimusaineistosta löytyviä havaintoja vasten. Ominaispiirteet näyttäytyvät eri tavoin keskeisinä eri tutkijoiden käyttämissä käsitteistöissä. Ilmiöille yhteistä on, että ne ovat tendenssejä, eivät tarkasti mitattavia suureita. ”Sounding natural is a matter of degree” (Abney (1995): 12).

Esittelen yhteisesiintymän ominaispiirteitä tämän alaluvun osioissa 2.2.1–2.2.2 ja vakiintumalle tyypillisiä ominaispiirteitä vastaavasti osiossa 2.2.3. Assosiaatioon liittyvät ominaispiirteet liittyvät molempiin, vaikkakin niillä on yhteistä rajapintaa yhteisesiintymien syntymekanismien kanssa ja ne ovat siinä mielessä tutkimuksen fokuksen ulkopuolella. Pohdin näitä ominaispiirteitä osiossa 2.2.4. Ajattelen, että yhteisesiintymä (fokus frekvenssissä) on kuin fyysinen todellisuus, konkreettinen, laskettavissa oleva suure. Vakiintuma (fokus merkityksessä) puolestaan edustaa abstraktia ominaisuutta ja on vaikeammin todennettavissa ja enemmän sidoksissa sekä teksti- että tilannekontekstiin. Yhteisesiintyminen tuottaa potentiaalisia, leksiikaalistuvia vakiintumia, samalla kun sanayhdistelmän osat deleksikaalistuvat ja menettävät itsenäistä merkitystään (Sinclair ja Carter (2004)). Assosiaatioon liittyvät ominaispiirteet liittyvät molempiin, vaikkakin niillä on tässä tutkimuksessa ainakin lähtökohtaisesti enemmän yhteistä rajapintaa intensionaalisen näkökulman ja yhteisesiintymien syntymekanismien kanssa ja ne ovat siinä mielessä tämän tutkimuksen fokuksen ulkopuolella.

Tutkimuksen näkökulma on distributionaalinen ja havainnon ensisijainen arviointikriteeri on esiintymistäajuus: tavoitteena on laskea esiintymiä ja arvioida ilmiön yleisyyttä aineistoissa niiden perusteella, siis kartoittaa yhteisesiintymiä. Havaintojen perusteella siirrytään tarkastelemaan, missä määrin niitä voidaan arvioida merkitykseen liittyvillä kriteereillä käytettävissä olevan aineiston perusteella. Merkitykseen liittyvät havainnot täydentävät siten määrällisiä havaintoja.

- Yhteisesiintymän distributionaaliin ominaisuuksiin liittyvät ominaispiirteet: toistuvuus/tuttuus/tyypillisuus/todennäköisyys (2.2.1)
- Yhteisesiintymän ja vakiintuman kokoonpanoon ja syntaksiin liittyvä ominaispiirre: rakenne/koostumus (2.2.2)
- Vakiintuman semanttinen ominaispiirre: merkitys/intensio (2.2.3)
- Vakiintuman psykologinivinen ominaispiirre: miellelyhtymä/assosiaatio (2.2.4)

Seuraavaksi tarkastelen ominaispiirteitä tarkemmin edellä mainitussa järjestyksessä. Kahteen ensimmäiseen ominaispiirteiden ryhmään liittyvät tutkimustulokset on esitelty luvussa 6 ja kahteen jälkimmäiseen luvussa 7.

2.2.1 Toistuvuus, tuttuus, tyypillisuus, todennäköisyys

Tässä alaluvussa käsittelem tämän tutkimuksen kannalta keskeisintä ominaispiirrettä, havainnon toistuvuutta. Lähtökohtana on ajatus siitä, että ilmiön tulee olla aineistossa toistuva, jotta siitä voi tehdä edes rajattuja johtopäätöksiä. Kääntäjän kannalta kuitenkin myös yksittäinen vastineen valinta voi muodostua oleelliseksi, jos se valikoituu oman erikoisalansa termityökaluun. Yksittäiset esiintymät voivat myös tuoda esiin kääntäjän tekemiä valintoja, jotka eivät ole peräisin apuvälineinä käytetyistä termipankeista tai konekäännöstyökalusta, jolloin niistä voi nähdä mahdollisia kollokaatioiden tuotaamiseen liittyviä syitä. Tämä ristiriita toistuvuuden ja yksittäisen esiintymän todistusvoiman välillä on ollut tutkimuksen ytimessä koko sen elinkaaren ajan.

Kollokaatio on abstrakti ilmiö, jonkinlainen toistuvuuteen ja tottumukseen perustuva sidos sanojen välillä ja se realisoituu konkreettisine yhteisesiintyminä puheessa ja tekstissä. Kielenpuhujat tunnistavat poikkeamat kieltään koskevista tottumuksista ja kielen käyttötavoista helposti, vaikkakin näiden poikkeamien tuottaminen tietoisesti tai arviointi on monesti vaikeaa. Esimerkiksi lähisyronyymejä käsittelevässä tutkimuksessaan Zaiu Inkpen ym. toteavat, että potentiaalisten kollokaatioiden (24 lähisyronyymiä ja 916 potentiaalista kollokaatiota) arviointi oli tutkimuksen kahdelle arvioijalle vaikea tehtävä. Arvioijat eivät saavuttaneet yksimielisyyttä

keskenään eivätkä olleet omissa arvioissaankaan täysin johdonmukaisia. (Inkpen ja Hirst (2002))

Sinclair kuvaa ilmiötä intuitioon perustuvaksi tunnistamiseksi, että jokin 'esiintyy usein' ja hän kattaa sen termillä **regular collocation**. Hänen mielestään voisi vaihtoehtoisesti puhua **tärkeästä** (important) tai **tutusta** (familiar) ilmiöstä, kun taas **merkitsevä** (significant) kuvaa tilastollista merkitsevyyttä (Sinclair et al. (2004)). Firthiläisen tradition puolella yhteisesiintyminen määritelläänkin yksittäisten leksikaalisten elementtien **tavanmukaiseksi** yhdessä esiintymiseksi: "habitual co-occurrence of individual lexical items" (Crystal (2003): 82). Englanninkielisten termien 'collocation' ja 'co-occurrence' käyttö on kirjavaa ja näistä jälkimmäisen avulla määritellään usein ensimmäinen, kuten Crystalin määritelmässä yllä. Omasa ajattelussani yhdistän Sinclairin **tuttuuden** käsitteen viestin vastaanottajan, kuulijan tai lukijan, reaktioihin kuulemaansa tai lukemaansa kohtaan, mikä ei ole tämän tutkimuksen menetelmin mitattavissa muutoin kuin oman tutkijan kielitajuni varassa (ks. osio 2.2.4).

Tieteen termipankkiin kirjatun määritelmän mukaan yhteisesiintymän osat esiintyvät **usein** yhdessä: yhteisesiintymä on "usein yhdessä esiintyvien sanojen muodostama kokonaisuus", josta esimerkkejä ovat *ajaa autoa, lehmä ammuu, rauhanomaisen rinnakkaiselo, illa berörd*.⁹ Arvioitaessa, onko jokin asia tavallinen aineistossa tai meitä ympäröivässä todellisuudessa, turvaudutaan toisinaan tutkijan tai kielenpuhujan intuitiiviseen käsitykseen siitä, mitä tarkoittaa 'usein' tai 'useammin kuin satumatlta'. 1980-luvun alkupuolella, jolloin John Sinclair tutki kahden kielellisen elementin muodostamien elementtien esiintymistajuuksia COBUILD-sanakirjatyötä varten, riitti hänen mukaansa kahden yhteisesiintymän löydös antamaan leksikografille aihetta tarkastella esiintymän relevanssia sanakirja-artikkeliksi. Käytettävissä olevat, tietokoneen avulla analysoitavat aineistot olivat tuolloin varsin suppeita (Sinclair (1996): 81). Sinclairin ja Teubertin välisessä keskustelussa (Sinclair et al. (2004)) Teubert sanoo: "...if we look at a list of highly significant bigrams, co-occurrences of two words, many of them seem not to inform us of anything. It was, after all, J. R. Firth who pointed out that we do have to measure statistical relevance against intuitive categorisation." (Sinclair et al. (2004): xxii). Firth siis arveli sanojen esiintymisen toistensa läheisyydessä olevan ennustettavissa, jolloin tämä läheisyys on myös numeerisesti esitettävissä. Useus (määritellyssä populaatiossa) on tarkasteltavissa esiintymiä laskemalla. Tämä on käsillä olevan tutkimuksen perustavaa laatua oleva lähtökohta.

Evertin tutkimusten distributionaalinen lähestymistapa (distributional approach, Evert (2005): 15) jatkaa neo-firthiläisen koulukunnan työtä leksikaalisten, empii-

⁹Tieteen termipankki 14.05.2016: Kielitiede:kollokaatio
(Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Kielitiede:kollokaatio>.)

risesti todennettavien, odotuksenmukaisten sanayhdistelmien (word combinations) parissa ”as empirical statements about the predictability of word combinations” (Evert (2008): 2), jossa komponenttien välisten suhteiden ennustettavuus ilmaistaan tilastollisen todennäköisyyden avulla. Sanat ulottavat **tilastollista vaikutusta** muihin sanoihin (Evert (2008)). Kollokaatio muodostuu sanojen yhteisesiintymisestä (cooccurrence) ja toteutuu kolmella tasolla. **Pintamuotojen** tasolla (surface cooccurrence) tutkitaan tekstistä lähellä toisiaan esiintyvien sanojen yhteisesiintymiä, **tekstin** tasolla (textual cooccurrence) virkkeen, lauseen, kappaleen tai dokumentin sisäisiä ja **syntaksin** tasolla (syntactic cooccurrence) tarkastellaan syntaktisessa suhteessa olevia yhteisesiintymiä (Evertin tapauksessa bigrammeja). (Evert (2008): 4–5)

Esimerkkinä syntaktisen tason tutkimuksesta Krenn ja Evert tutkivat muun muassa saksan verbilausekkeita ja kuvakielisiä ilmauksia määrittelemällä etukäteen syntaktisesti yhteisesiintyjien ryhmään hyväksyttävät yhdistelmät (Krenn ja Evert (2001)). Heidän tavoitteenaan oli tutkia, mikä osuus tilastollisten menetelmien tarjoamista mahdollisista yhteisesiintymistä on todellisia yhteisesiintymiä. Tutkimuksessa keskityttiin selvittämään, miten suuri määrä tilastollisin menetelmin löydettyistä potentiaalisista yhteisesiintymistä osoittautui ihmisten tekemän arvion perusteella todelliseksi tai oikeiksi yhteisesiintymiksi. Potentiaalisia yhteisesiintymiä heidän tutkimuksessaan ovat figuratiiviset ilmaukset sekä PP-verbikonstruktiot, joilla on myös kuvaannollinen merkitys, esimerkiksi ”*am Herzen liegen*” (’olla sydäntä lähellä’, ’*maata/olla/sijaita sydämessä’). Heidän havaintonsa tukevat merkityksestä lähtevän tarkastelun erottamista puhtaasti tilastollisiin havaintoihin perustuvasta menetelmästä. Tutkimuksessani olen valinnut syntaktisen kriteerin lähtökohdaksi, tutkin verbin ja objektin asemassa olevia elementtejä ja niiden välistä kollokaation muodostavaa suhdetta ja elementtien variaatiota. Tutkimuskohteena olevat tutkimuskohteen mahdollisina yhteisesiintyjinä tutkitut verbit näyttäytyvät synonyymisinä. Kollokaatioanalyysin avulla voidaan tarkastella kääntäjän valintoja ja niiden vaikutuksia tutkimuskohteen merkityksen muodostumiseen.

Toistuvuutta arvioidaan tässä tutkimuksessa tilastollisten menetelmien tarjoamien tunnuslukujen kautta. Tilastollisen menetelmän valinta riippuu siten kollokaation määritelmästä ja tavoitteista. Menetelmillä on luonnollisesti eri lähtökohdat ja oletukset tutkittavasta ilmiöstä. Ne tuottavat samasta tutkimusaineistosta erilaisia lopputuloksia. Esimerkiksi t-testi ”...tends to favour collocates that are high frequency function words..” ja ”Pointwise mutual information ... favours rare content words ...”(Hanks (2012): 57, Stubbs (1996)). MI-testi olettaa kollokaation elementtien keskinäistä vastavuoroisuutta, se laskee molempien esiintymiset erikseen ja yhdessä ja merkitsevyys määräytyy sen perusteella, miten usein yhteisesiintyminen toteutuu molempien elementtien kohdalla. Kaksi harvinaista sanaa, jotka esiintyvät poikkeuksetta yhdessä, nousevat listalla korkeimmalle sijalle ja menetelmä löytää

aineistosta siten pitkälle vakiintuneita kollokaatioita. Clear (1993) ilmaisee asian näin: ”In general, the collocates which are identified by MI are a rather motley and peculiar set” Clear (1993): 279. Adam Kilgarriff tarkastelee tilastollisten, nollahypoteesin testaukseen perustuvien menetelmien hyödyntämistä korpuslingvistiikassa usein kriittisesti, ja esittää, että nollahypoteesi (kielelliset yksiköt ovat aineistossa riippumattomia toisistaan) voidaan riittävän laajoilla aineistoilla lähes poikkeuksetta kumota. Vaikka voidaan osoittaa kielellisten ilmiöiden välisen suhteen olevan satunnainen (non-random), ei voida suoraan päätellä, että kyseiset ilmiöt eivät ole sattumanvaraisia (arbitrary). (Kilgarriff (2005))

Erikoiskielten ja terminologian merkityksen tutkimuksessa usein esiintyminen tai tilastollinen merkitsevyys ei ole samalla tavoin oleellista, koska termin status ja merkityssisältö perustuvat kieltä käyttävien sopimukseen. Yhdyssana tai moniosainen termi voi esiintyä erikoiskielen teksteissä usein, mutta niin ei välttämättä ole. Esiintymistajuus ei siten voi olla termin kriteeri, eikä sen harvinaisuus aineistossa poista termiltä sen statusta. (Pearson (1998): 123) Voidaan siis olettaa, että erikoiskielinen aineisto on termien suhteen tiivistä ja myös yleiskieliseltä aluksi vaikuttava kielellinen elementti voi olla termi.

Kollokaatio voidaan nähdä myös korpusten käsittelyn menetelmänä: se on ”The final method in the corpus toolbox” (Rayson (2015): 41). Aineistosta havaitut kollokaatit auttavat (haku)sanan merkityksen määrittämisessä. Menetelmä tarkoittaa Raysonin mukaan laskemista, usein tilastollisen merkitsevyyden määrittämistä, konkordansseja ja avainsana-analyysia, mutta kollokaation tunnistamisen menetelmä ei ole selkeästi määriteltävissä, etenkin kun tutkimusasetelman valinnoilla ja muutoksilla (esimerkiksi tutkimusikkunan koko) on suuri vaikutus tuloksiin.

Todennäköisyys terminä on tässä tutkimuksessa pyritty varaamaan tulevaisuuden tapahtumille, ei jo tapahtuneille ja tekstistä havaittavissa oleville. Siten tekstistä saadut havainnot ja tunnusluvut ovat todellisia, jo tapahtuneita yhteisesiintymän toteutumia, instanttaatioita. Esiintymistajuuden avulla kuvataan kielen käytössä havaittua toistuvuutta (regularities of performance) ja todennäköisyydellä viitataan havaintojen perusteella tehtävään arvioon ilmiön tapahtumisesta tulevaisuudessa tai toisessa kontekstissa (regularities in the system). Toury ilmaisee asian seuraavasti: ”Put in slightly different terms, one could perhaps say that frequency applies first and foremost to things *past*, whereas probability makes a claim for validity in the *future*” (Toury (2004): 19–20).

Kollokaation ominaispiirteitä on määritelty etukäteen, mutta ne ovat koko tutkimuksen elinkaaren ajan muuttuva ja rajaamaton osa-alue, jonka tarkentaminen on tutkimuksen yhtenä tavoitteena. Tilastollisten menetelmien hyödyntäminen edellyttää kuitenkin niiden käsitteistön, taustamuuttujien ja laskentalogiikan tuntemusta. Tilastollisia menetelmiä on sen vuoksi käytetty tutkittavien kielellisten elementtien

valinnan perusteena tutkimuksen alkuvaiheessa sekä myöhemmin aineistoista tehtyjen laskelmien vertaamiseen toisiinsa siinä määrin kuin edellä mainittu tuntemus on osoittautunut mahdolliseksi saavuttaa. Tilastollisesta analyysistä aineistojen vertaamisessa ks. alaluku 5.5.

2.2.2 Rakenne

Tarkastelen tässä osiossa yhteisesiintymän/vakiintuman rakenteeseen liittyviä seikkoja: minkälaisista elementeistä se katsotaan yleensä koostuvaksi, miten kaukana nämä elementit sijaitsevat toisistaan kieliaineistossa ja miten monesta elementistä yhteisesiintymä voi koostua. Rakenteellisia ominaispiirteitä ovat siten yhteisesiintymän **saneiden lukumäärä**, osien **fyysinen läheisyys**, **peräkkäisyys** ja **sisäkkäisyys**. Kieliopilliseen rakenteeseen kiinnitän selvästi vähemmän huomiota, koska se on suomenkielisessä aineistossa itsessään laaja tutkimuskohde. Joitakin syntaksiin liittyviä ominaispiirteitä on valittu tarkasteltavaksi sen perusteella, miten niistä on tutkimuskirjallisuudessa esitetty havaintoja.

Yhteisesiintymistä kartoittavissa tutkimuksissa mielletään yhteisesiintymän osa tavallisimmin **sanaksi**. Osia voivat kuitenkin olla myös mitkä tahansa muut kielellinen elementit, esimerkiksi symbolit, lyhenteet ja välimerkit (ks. 2.2.3). Tutkimukseni laskettavissa oleva elementti on **sane**. Se käsittää kaikki välilyönnillä toisistaan erotetut ortografiset elementit. Symboleilla ja lyhenteillä on sama tehtävä viestinnässä kuin sanoilla. Säädöskielessä tyypillinen symboli on pykälänmerkki ja tyypillinen lyhenne on lakia tarkoittava L.

Yhteisesiintymä koostuu osista, joita on englanninkielisen kirjallisuuden perusteella vähintään kaksi. Sinclair määritteli yhteisesiintymän vuonna 1991 seuraavasti: "The co-occurrence of two or more words within a short space of each other in a text." (Sinclair (1991): 170). Yhteisesiintymän osia ovat hakusana ja yhteisesiintyjä(t). Jos analyysi tehdään manuaalisesti, yhteisesiintymän osien lukumäärää on kuitenkin käytännössä rajoitettava, mikä saattaa selittää Sinclairin näkemyksen. Nykyisin tavanomainen laajojen teksti- tai puhekorpuksen käsittely ja analyysi ei vielä 1990-luvun alussa ollut tietoteknisesti mahdollista. Sinclair päätyi suosittelemaan ikkunaa, jossa hakusanan ympäriltä tutkitaan molemmilta puolilta neljän yksikön verran tekstiyhteyttä (Sinclair ja Carter (2004)). Heid rajaa tutkimuksessaan yhteisesiintymän osien lukumäärän tarkasti kahdeksi ja määrittelee useamman kuin kahden sanan yhteisesiintymät kuuluvaksi yhteisesiintymisen ulkopuolelle muun muassa monisanaisiksi elementeiksi/yksiköiksi tai idiomeiksi. Terminologian näkökulmasta ainakin toisen yhteisesiintymän osista on oltava termi, jonka käsite on määritelty. (Heid (2001): 788–789). Hallidayn mukaan yhteisesiintymäinen toimii kaikilla tekstin tasoilla, jolloin sen reaalitumana yhteisesiintyjä tai vakiintuma voi ylittää virke- ja

kappalerajat (Halliday ja Hasan (1985)).

Myös suomea käsittelevissä tutkimuksissa yhteisesiintymän osien lukumäärä on tavallisesti kaksi, muun muassa Jantunen (2004), Uosukainen (2006). Jantunen tarkastelee sanojen vieruskollokaatteja, toisin sanoen heti hakusanojen vieressä sijaitsevia kollokaatteja ja niiden synonymiaa suomeen käännettyssä aineistossa ja alun perin suomeksi kirjoitetuissa teksteissä (Jantunen (2004)). Arppe totesi neljän verbin synonymiaa käsittelevässä tutkimuksessaan, että suurin osa yhteisesiintyjistä löytyy +/- 5 saneen kokoisen ikkunan sisällä (Arppe (2008)). Viime aikoina on myös laajempia yksiköitä hyödynnetty tutkimuksessa. Aino Piehl etsi EU-direktiivien revisioista 3-, 4- ja 5-grammeja havainnoidessaan lainvalmistelijoiden tekemiä muutoksia. Piehl toteaa, että n-grammit eivät sellaisenaan tarjoa riittävästi materiaalia tehtyjen muutosten tunnistamista varten, vaan niitä on tutkittava kontekstissaan. 3- ja 4-grammit haettiin aineistosta identtisinä, mutta 5-grammeissa sallittiin vaihtelua muun muassa pronominien ja välimerkkien osalta. (Piehl (2013): 165–166)

Säädöskielen virkkeet ja lauseet ovat usein pitkiä ja voidaan olettaa, että yhteisesiintymän osat voivat sijaita tekstissä kaukana toisistaan. Jos halutaan sallia sekä iso tutkimusikkuna että yhteisesiintymän osien suuri määrä, tarvitaan dependenssijäsennettyä aineistoa tai päädytään raskaisiin tilastollisiin ajoihin ja isoihin tulosjoukkoihin, joiden käsittelyssä tarvitaan tämän tutkimuksen ulottumattomissa olevia menetelmiä.

Vaikka yhteisesiintyjien lukumäärää ei olisikaan rajoitettu, on tämän tutkimuksen menetelmillä useamman kuin kahden saneen muodostavia yhteisesiintymiä käytännössä mahdollista tutkia vain hyvin pienistä aineistoista käsityön suuren määrän vuoksi. Toinen ratkaiseva päätös on tutkimusikkunan koko, joka määrää tutkittavien saneiden väliin jäävien saneiden maksimimäärän. Tutkin ennalta määrättyä substantiivisia ja sen kanssa esiintyviä etukäteen määriteltyjä verbejä, mutta laadullinen analyysi mahdollistaa myös näiden lisäksi samassa syntaktisessa rakenteessa esiintyvien elementtien tarkastelun: voin esimerkiksi tarkastella substantiivisia määrittäviä adjektiiveja ja pyrkiä tunnistamaan kahta elementtiä laajempia kollokaatioita.

Yhdyssanat erikoistapauksena

Kielissä, joissa yhdyssanojen muodostus on produktiivista (mm. suomen, ruotsin ja saksan kielissä), voivat ortografisesti alun perin erikseen kirjoitetut sanat usein toisessaan kiteytyä yhdyssanoiksi (esim. ruotsin *så som* vs. *såsom*). Yhteen ja erikseen kirjoitetut versiot saattavat myös elää kielessä rinnakkain. Niillä voi kuitenkin olla omia konnotaatioita tai merkityksiä, toisinaan merkitysero voi olla referentiaalisen merkityksen sijaan pragmaattinen. Esimerkiksi suomen *pöydän jalka* viittaa tiettyyn tarkoitteeseen ja sen valitseminen korvaa osaltaan suomesta puuttuvaa ilmiäntä.

määräistä artikkelia, kun taas *pöydänjalka* viittaa geneeriseen tarkoitteeseen. ”Yhdyssana on yleensä merkitykseltään kiteytyneempi, leksikaalistuneempi kuin kaksisanainen vastineensa” (Iso suomen kielioppi, Hakulinen et al. (2004): §221). Myös englannin kielessä osien kirjoittaminen yhteen tuottaa toisinaan merkityksen muutoksen ja tuo tarkoitteen yleiseltä tasolta yksityiseksi, esimerkiksi *black bird* (värittään musta lintu) – *blackbird* (lintulajin nimi, mustarastas) (Sinclair ja Carter (2004): 26).

Tarkastelen yhdyssanoja tässä tutkimuksessa potentiaalisina yhteisesiintyminä kiinnittäen huomiota siihen, miten ne näyttäisivät käyttäytyvän yhteisesiintymän osana. Oletan, että suomen kielessä yhdyssana on yhteisesiintymän ja vakiintuman erikoistapaus. Tutkin yhdyssanoja myös itsenäisinä kahden komponentin, edusosan ja määriteosan muodostamina yhteisesiintyminä.

Syntagma, jono, sekvenssi, osien järjestys, läheisyys, vierekkäisyys

Firth kuvaa kielen organisoitumisen mekanismeja, syntagmaa¹⁰ ja paradigmaa artikkelissaan *A Synopsis of Linguistic Theory, 1930–55* (Palmer (1968): 168–205), Firthin sanoin: ”Structure consists of elements in interior syntagmatic relation and there elements have their place in an order of mutual expectancy.” (Palmer (1968): 200). Systemit puolestaan realisoituvat paradigmaattisella tasolla: ”Systems of commutable terms or units are set up to state the paradigmatic values of the elements” (mt.). Kumpikaan niistä ei viittaa kielen konkreettisiin elementteihin. Ne ovat abstrahoituja kategorioita, joiden avulla kielellisten elementtien välisiä suhteita voidaan kuvata (Palmer (1968): 186). Yhteisesiintyminen on eräs kielellisten elementtien välisistä suhteista. Syntagma on jono¹¹ ja sekvenssillä¹² tarkoitetaan peräkkäisyyttä.

¹⁰Syntagma; Määritelmä: peräkkäisten kielellisten yksiköiden muodostama suurempi yksikkö, jossa nämä muodot ovat sekventiaalisessa suhteessa toisiinsa, Selite: Syntagma on aina kahden tai useamman yksikön, esim. sanojen, tavujen tai foneemien muodostama sekvenssi. Esimerkiksi sana *susi* on nelifoneeminen syntagma, lause *Susi ulvoi* kaksisanainen syntagma. Peräkkäisten sanojen syntagmassa yksiköitä erottelee esimerkiksi merkitys tai sanapaino. (Lähde: Tieteen termipankki 2.03.2016: Kielitiede:syntagma.)

¹¹Syntagma, Lisätiedot: Syntagman käsitteen luoja, Ferdinand de Saussuren mukaan olennaista on sen muodostamien yksiköiden välinen riippuvuus: ”There are always larger units, composed of smaller units, with a relation of interdependence holding between both”. Saussure kutsui syntagmaa joskus ketjuksi; F. Karlsson on käyttänyt siitä sanaa *jono*. (Lähde: Tieteen termipankki 2.03.2016: Kielitiede:syntagma.)

¹²Syntagmaattinen suhde; Määritelmä: kielenyksiköiden sekventiaalinen suhde toisiin yksiköihin, joiden kanssa se esiintyy peräkkäin samassa jaksossa tai kontekstissa. Selite: Esimerkiksi sana toimii yksikkönä syntagmaattisessa suhteessa, jonka rakentumista säätelee syntaksi, foneemit puolestaan muodostavat syntagmaattisesti jonoja liittyessään toisiinsa fonotaktisin periaattein. Lisätiedot: Termi on peräisin Ferdinand de Saussurelta. Roman Jakobsonin poeettisen kielen teoreemassa termi yhdistännän akseli (suom. P. Leino 1982) tarkoittaa syntagmaattista ulottuvuutta. (Lähde: Tieteen

Esipuheessaan Palmer (1968) huomauttaa kuitenkin, että Firth ei tehnyt eroa kielellisten elementtien sekvenssin eli sarjan (sequence) ja järjestyksen (order) välillä (Palmer (1968), Firth (1957)).

Lähietäisyys ja välitön läheisyys ovat suhteellisia, rekisteriin sidoksissa olevia käsitteitä. Ne mielletään useimmiten lauseen tai virkkeen sisäisiksi, kuten Sinclair toteaa vuonna 2004 ilmestyneessä teoksessaan *Trust the Text*: ”I would like to suggest, as an alternative, that the most important thing is what is happening in the current sentence.” (Sinclair et al. (2004): 13). Englanninkielisessä tutkimuksessa yhteisesiintymien määrittelyä usein lähietäisyyden käsitteen avulla.

Yhteisesiintymän osien ei siten tarvitse sijaita jonossa vierekkäin tai edes toistensa läheisyydessä. Tutkimukseni näkökulmasta tämä tarkoittaa sitä, että merkityksen analyysin kannalta ei ole ratkaisevaa, missä kohtaa syntagmaattisessa rakenteessa kollokaation jäsenet sijaitsevat. Ne voivat sijaita sanojen muodostamassa jonossa vierekkäin tai niiden välissä voi olla muita sanoja. Vaikka Sinclair ei rajoita itse ilmiön ulottuvuutta, hän määrittelee englannin kieltä koskevien havaintojensa perusteella noodin ja yhteisesiintyjän (kollokaatin) väliseksi etäisyydeksi neljä sanaa (kollokaatti mukaan lukien) (Sinclair ja Carter (2004): 141) ja kuten edellä jo mainittiin, Arppe (2008) tuli suomen osalta alustavaan tulokseen viiden sanan etäisyydestä. Halliday korostaa kollokaation koheesiota luovaa merkitystä, joka ulottuu virke- ja kappalerajojen ulkopuolelle tekstin tasolle ja ilmenee tekstissä erilaisina olomuotoina, esimerkiksi synonymiana tai lähisynonymiana, komplementaarina rakenteina tai tunnistettavan kokonaisuuden osana (Halliday (1978)).

Vakiolausekkeissa sanojen järjestys on toisinaan tärkeä asia, mutta näin ei välttämättä aina ole. Konventionaalisten fraasien sanajärjestys on tärkeää pitää samana, erityisesti ikonisissa käyttöyhteyksissä. Esimerkiksi vihkipuheessa morsian ja sulhanen vihitään avioliittoon ”mieheksi ja vaimoksi”, ei koskaan toisin päin.

? % totean	? % A:n ja B:n	? % näin	0 % vaimoksi ja 0 % mieheksi
100 % julistan	? % heidät	? % täten	100 % mieheksi ja 100 % vaimoksi

Kuva 2.1: Syntagma ja valistunut arvaus paradigmaattisesta vaihtelusta

Tutkimuskohteena ja hakusanana on substantiivi, jonka tekstiyhteydestä tutkin tarkemmin kuuden verbin esiintymiä. Sinclairin tutkimusten esimerkin mukaisesti tarkastelen hakusanaa tekstiyhteydessä, jonka laajuus on valittu menetelmän toteuttamiskelpoisuuden näkökulmasta siten, että tulosjoukot on vielä mahdollista tarkistaa manuaalisesti. Yllä mainittu viiden sanan ikkuna hakusanan molemmiin

puolin täyttää tämän ehdon, tosin tässä tutkimuksessa elementit ovat sanojen sijaan saneita. Tutkimusikkunan laajuudeksi valikoitui siten 11 sanetta. Hakusana voi sijaita jonossa ennen sen yhteisesiintyjää tai sen jälkeen ja niiden välillä voi olla muita saneita tai merkkejä, mukaan lukien välimerkit. Yhteisesiintymä voi koostua kahdesta tai useammasta saneesta, jotka voivat esiintyä vierekkäin, mutta niiden välissä voi olla myös muita leksikaalisia elementtejä tai saneita. Poikkeuksen ikkunan kokoon tekee kolmas verrannaisaineisto, jossa tarkastelualueena on koko virke. Tutkimusaineistot on kuvattu luvussa 4 ja tutkimusmenetelmä luvussa 5.

Kiinteys ja sisäkkäisyys

Heid kuvaa sisäkkäisyyttä seuraavasti: ”The elements of collocations can be collocations themselves” (Heid (2001): 789). Yhteisesiintymät voivat siten sijaita myös toisten yhteisesiintymien sisällä. Heid käyttää ilmiöstä nimitystä ”verbin mini-paradigma” (Heid (2001): 789) ja se kattaa esimerkiksi tilanteen, jossa yhteisesiintymän rakenne on lähtökohtaisesti substantiivin ja verbin yhdistelmä N+V, johon kuitenkin liittyy myös usein yhteisesiintymän osana esiintyvä adjektiivi. Heidin esimerkki saksan kielestä on *unbeschränkte Gültigkeit haben* (esim. ’omata’ esim. ’pätevyys’). Mikä adjektiivi sopisi Suomen säädöskielessä täydentämään lauseen? Asiaa tarkemmin tutkittaessa huomataan, että EU-kontekstissa sanayhdistelmän merkitys on ’olla rajoittamattomasti voimassa’: *Form I-551 Alien registration receipt card (no expiry date)*¹³ / *Form I-551 Alien registration receipt card (Wiedereinreisedokument) (unbeschränkte Gültigkeit)*¹⁴.

Sisäkkäisiä yhteisesiintymiä tutkinut Seretan (2011) kutsuu kahdesta sanasta koostuvaa yhteisesiintymää binääriseksi kollokaatioksi (binary collocation), kun taas kompleksinen kollokaatio (complex collocation) koostuu useammasta kuin kahdesta sanasta, esimerkiksi kun kahden sanan muodostamaan vakiintumaan liittyy adjektiivi: *draw a clear distinction* (Seretan (2011): 5). Omassa tutkimuksessani oletan, että kollokaatio voi koostua useammasta kuin yhdestä elementistä ja käytän näissä tapauksissa nimityksiä **ensimmäisen/toisen/kolmannen** jne. **asteen yhteisesiintymä** tai **monisanainen yhteisesiintymä/vakiintuma**.

¹³VERORDNUNG (EG) Nr. 810/2009 DES EUROPÄISCHEN PARLAMENTS UND DES RATES

¹⁴REGULATION (EC) No 810/2009 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

Syntaksi

Yhteisesiintymä voidaan jo ennen aineistosta tehtäviä hakuja varten määritellä syntaktisista kriteereistä käsin, jos käytettävissä on jäsennetty aineisto. Menetelmä edustaa ajattelutapaa, jossa tutkijan intuitioon tai jo tehtyyn tutkimukseen perustuen määritellään ennakoita, mihin sanaluokkaan tai lauseenjäsenen luokkaan hakusana ja/tai yhteisesiintyjä kuuluu. Jos esimerkiksi ajatellaan, että yhteisesiintymän pääsana tekstissä on aina substantiivi ja yhteisesiintyjä useimmiten toinen substantiivi, verbi tai adjektiivi, voidaan käsitelystä aineistosta etsiä vain ne edustajat, joissa hakusanan kanssa esiintyy halutun tyyppisiä ehdokkaita. Näin tulosjoukko saadaan riittävän pieneksi manuaalista käsittelyä varten. Erikoiskielissä, myös säädöskielessä, juuri substantiivit toimivat yhteisesiintymien ytiminä ja termipankkien artikkeleina ne ovat siten luontevia valintoja hakusanoiksi. Erikois- ja ammattikielissä yhteisesiintymien mekanismina toimii kuitenkin myös verbisidonnaisissa ja adjektiivisidonnaisissa yhteisesiintymissä (Helin (1998)) ja englantia koskevissa tutkimuksissa myös partikkeleita on tunnistettu yhteisesiintymän ensisijaiseksi komponentiksi (Sinclair (1991)).

Yhteenveto: rakenne

Tarkastelun kohteena tässä tutkimuksessa on etukäteen määritelty substantiivi, jota pidän lähtökohtaisesti yhteisesiintyjän pääsanana. Sen potentiaalisina yhteisesiintyjinä tarkastelen kuutta verbiä ja käytän tutkimusikkunana hakusanan lisäksi viiden saneen pituista jonoa tutkimusaineistossa hakusanan eli tutkittavan substantiivin molemmin puolin. Potentiaaliset yhteisesiintyjät voivat sijaita kummalla puolella hakusanaa tahansa. Tarkastelen joitakin yhdyssanoja erikseen huolimatta siitä, että ne eivät toteuta kahden ortografisen saneen lähtökohtaista kriteeriä. Tarkastelen myös verbikantaisia substantiiveja verbien edustajina, vaikka tässä valinta tehdään merkitykseen eikä sananmuotoon perustuen. Valinta tuottaa tutkimukseen tietyn ristiriidan, kun toisaalta ei ole tarkoituksena tutkia vakiintumia tai muutoin merkitykseen perustuvia sanayhdistelmiä. Valinta on tehty tutkimusaineiston tuloksia käsin analysoitaessa ja verbikantaisten substantiivien havaitseminen on eräs tutkimuksen mukanaan tuomista yllättävistä valintatilanteista. Palaan tähän kysymykseen tutkimuksen johtopäätösluvussa 8.

Seuraavassa alaluvussa tarkastelen kollokaation merkitykseen liittyviä ominaispiirteitä ja pohdin tämän tutkimuksen mahdollisuuksia tarkastella niitä tutkimusmenetelmän puitteissa.

2.2.3 Merkitys/intensio

Tämä alaluku keskittyy merkityksen tarkasteluun kollokaation muodostumisessa kollokaation syntymekanismina. Tarkastelen ensin kontekstin roolia merkitystä rajaavana tekijänä ja pohdin lyhyesti sanan perusmerkitystä ja yhteisesiintymisen vaikutusta merkityksen muodostumiseen. Todettakoon, että monipuolinen aiheen käsittely semantiikan puolella on tämän tutkimuksen ulottumattomissa, joskin se olisi kääntäjän kannalta luonnollisestikin hyödyllistä. Merkitykseen liittyvät käsitteet läpinäkyvyys ja kompositionaalisuus käsitellään lyhyesti, samoin systeemin ja paradigman käsitteet sekä vaihdettavuuteen ja lähisynonymiaan liittyvät ominaispiirteet, jotka tulevat tässä tutkimuksessa tarkastelluiksi vakiolausekkeiden komponenttien vaihdettavuuden havainnoimisen kautta. Lopuksi mainitsen semanttisen preferenssin ja prosodian käsitteet, joiden tarkastelu on jossakin määrin mahdollista aineistojen laajuuden asettamissa rajoissa.

Firthin sanoin, ”meaning is contextually determined” ja ”collocation is the tool for defining this contextual meaning” (Palmer (1968): 5). Konteksti luo, rajaa ja määrittää merkitystä. Merkitys vivahteineen muodostuu ilmaisussa mukana olevien kielellisten elementtien yhteiselistä käyttöyhteydessään ja rakentuu paitsi tottumusten ja konventioiden noudattamisesta, myös siitä kontrastista, joka syntyy kieltä käytettäessä tottumusten ja konventioiden vastaisesti (Kenny (2001)). Kielen kategoriat eivät ole tarkkarajaisia ja selkeitä. Ne ovat häilyviä ja horjuvia ja ilmenevät päällekkäisinä ja limittäisinäkin tasoina (Palmer (1968): 47, Shore (2004), Sinclair et al. (2004): 18). Sinclair näkee merkityksen muovautuvan leksikaalisten ja kieliopillisten elementtien välisten suhteiden kautta, jotka ovat probabilistisia ja joustavia, eivät tarkka-rajaisia ja pysyviä (Hanks (2008)).

Firth korosti merkitystä kielitieteen keskeisenä tutkimuskohteena (Honeybone (2005), Fellbaum (2011)). Tilannekontekstin konkreettiset ja abstraktit piirteet kokonaisuudessaan (mm. fyysinen ympäristö, aihe ja puhujat puhetilanteessa tai kielen käyttötilanteessa) rajaavat, vahvistavat ja tarkentavat kielellisten elementtien merkityksiä. Myötäesiintyminen on abstrakti sidos leksikaalisen elementin ja sen kanssa samassa merkitysyksikössä (kielellisessä yksikössä) esiintyvien toisten leksikaalisten elementtien välillä. Kielen puhuja hallitsee tämän sidoksen tiedostamattaan, automaattisesti. Yhteisesiintymä on siten osa kontekstia määrittämässä muiden kielellistä merkitystä kantavien yksiköiden valintaa. Miellämme leksikaaliset merkitystä kantavat yksiköt useimmiten sanoiksi, mutta ne voivat olla myös symboleja tai merkkejä.

Firthin näkökulma täydentää generativistista teoriaa kielestä ja se on tarjonnut kielen hahmottamiseen toisenlaisia painotuksia. Firth piti Saussuren ja Chomskyn näkemyksiä kielen kahtiajaoista kieleen (*langue*) ja puhuntaan (*parole*) ja toisaal-

ta kompetenssiin (competence) ja performanssiin (performance) harhaanjohtavina, koska ne vievät huomiota pois kontekstuaalisen merkityksen tutkimuksesta. Toisin kuin Firth, Chomsky asetti keskiöön kielen puhujan kompetenssin ja hylkäsi kielelliset autenttiset tuotokset epätäydellisinä ja siksi mielenkiinnottomina. Tilastollinen näkökulma kieleen on silloin hyödytön, koska kieli ei ole tilastollisilla menetelmillä mitattavissa olevan jatkumo vaan pikemminkin sarja joko – tai -valintoja. (Abney (1995): 4–5)

Mikä sitten on sanan perusmerkitys? Kontekstuaalisen teorian mukaan sanoilla on merkitysvivahteita, jotka aktivoituvat tilannekontekstin perusteella. On siis mahdollista, että tietyissä tilanteissa aktivoituu vain sanan jokin sanaluokkavariantti, pintamuoto tai taivutusmuoto, jolla on siten oma erityinen kyseisessä tilannekontekstissa aktivoituva merkityskenttensä, joka poikkeaa sanan muiden taivutusmuotojen tai sanaluokkien vastaavista. Sinclair esittääkin, että eri pintamuodoilla voi hyvinkin olla omat ainutkertaiset merkityksensä ja epäilee, ettei perusmuotoistaminen (ks. luku 5) korpuslingvistisessä tutkimuksessa ole välttämättä järkevä toimenpide (Sinclair et al. (2004): 17). Perusmuotoistaminen on agglutinoivaa kieltä tutkivana itselleni ainoa vaihtoehto tutkimusmentelmäni ja aineistojeni asettamien reunaehtojen vuoksi, vaikka kysymys pintamuotoisten saneiden yhteisesiintyjien jakaumista on mielenkiintoinen ja kääntäjänkin kannalta varmasti kiinnostava. Kysymys ei kuitenkaan ole tämän tutkimuksen fokuksessa. Sinclairin esimerkin verbi *move* ja vastaavat substantiivit *move* ja *movement* ovat useimmiten konkreettisia, kun taas adjektiivin *moving* merkitys on hänen mukaansa huomattavasti harvemmin konkreettinen kuin kuvaannollinen. Saman lemman eri sanaluokkavariantit voivat myös edustaa eri merkityksiä. Polyseemisten sanojen eri merkitykset eivät kielen tuottamisessa sekoitu (Sinclair et al. (2004): 18). Kysymys merkityksen ja sanaluokan välisen korrelaation murtumisesta realisoituu omassa aineistossani tutkittavana olevien substantiivien verbikantaisina substantiiveina, jotka on sisällytetty laskelmiin.

Kollokaatio on jaettua merkitystä tuottava mekanismi (Partington (1993): 186), (Sinclair et al. (2004): 19): ”Shared meaning is clearly at the heart of the principle of collocation. Very often, in the course of the on-line production of language, one word or group of words almost automatically ‘calls up’ another specific word or phrase, or at least, constrains the speaker to the choice of one of a limited set of possibilities.” (Partington (1993): 186). Vielä ei ole käytettävissä keinoja erotella, miten eri merkitysyksikön osat tuottavat yhteisen merkityksen, missä se alkaa ja mihin loppuu, kuten Partington totesi jo vuonna 1993: ”It is often impossible in the present state of our knowledge to say exactly where the realization of that meaning starts and stops...” (Partington (1993): 186). Oman tutkimukseni keinot asian tarkasteluun ovat rajalliset, ja pyrin todentamaan merkityksen muodostumista tutkimusaineistossa esille tulevien havaintojen osalta, en niinkään sellaisia etsien.

Evertin luokittelussa intensionaalinen lähestymistapa (intensional approach) on fraseologian tutkimuksen osa-alue, jossa kollokaatio muodostuu merkityksen perusteella (Evert (2005): 16). Keskiössä ovat leksikaalisin perustein määräytyvät sanayhdistelmät ”semi-compositional and lexically determined word combinations” (Evert (2008): 2). Näistä sanayhdistelmistä käytetään nimitystä moniosainen ilmaus, (MWE, multiword expression). Toisaalta moniosaisia ilmauksia tutkineet Baldwin ym. luokittelevat kollokaation ilmiönä näiden moniosaisien ilmiöiden alakäsitteeksi (proper subset) (Baldwin ja Kim (2010): 11). Moniosainen ilmaus on heidän määrittelemänään vähintään kahdesta komponentista leksikaalisin, syntaktisin, semanttisin, pragmaattisin tai tilastollisin kriteerein rakentuva idiomaattinen sanayhdistelmä (Baldwin ja Kim (2010): 4–7). Toisin kuin omassa tutkimuksessani, Baldwinin ym. tulkinnan mukaan yhdyssana yhteen kirjoitettuna ei kuulu moniosaisiin ilmauksiin, joihin englanninkieliset erikseen kirjoitetut yhdyssanat sisällytetään. Heidän määritelmänsä on siten kielikohtainen. Tämän tutkimuksen tavoitteena on pohtia ilmiötä sillä tavoin universaalina, että määritelmiä voisi hyödyntää yli kielirajojen ja riippumatta ortografisista rajoituksista.

Fraseologisen kollokaatiotutkimuksen yhtenä tavoitteena on määritellä semikompositionaalisia sanaliittoja tai merkityksen perusteella muodostettuja leksikaalisesti määräytyviä sanayhdistelmiä, esimerkiksi *heavy smoker*, jossa *heavy* on englannissa ainoa sallittu substantiivin määrite merkityksessä ’kova tupakoitsija’. Toisinaan sanan tavallisin merkitys sanayhdistelmässä on kaukana sen merkityksestä kollokaation osana, kuten esimerkissä *school of fish* (ei **group*, **swarm* tai **flock*) ’kalaparvi’. Toisinaan taas valinta tehdään arbitraarisempien leksikaaliset rajoitteiden vallitessa, kuten esimerkissä *brush teeth* (ei **scrub teeth*) tai joilla on vahvoja konnotaatioita (*red rose*). Moniosaisen vakiintuman määrittelemisen yhdeksi merkitysyksiköksi on hankalaa ja sanakirjoissa tyydytään usein kuvaamaan sanayhdistelmät listaamalla niiden elementit erillisinä hakusanoina ja sanayhdistelmät lisätietoina. (Evert (2005): 3)

Evertin tutkimuksessa keskeisiä ovat ilmauksen kompositionaalisuus ja läpinäkyvyys, ja sanayhdistelmien ajatellaan sijoittuvan jatkumolle läpinäkymättömistä/ei-kompositionaalisista idiomeista leksikaalisten kollokaatioiden kautta sanayhdistelmiin, joiden rajoitukset ovat mielivaltaisia. Moniosaisilla ilmauksilla on usein idiosynkraattisia semanttisia ja/tai syntaktisia ominaisuuksia, mikä tekee niistä haastavan osa-alueen luonnollisen kielen käsittelyssä. Tämän tyyppin kollokaatioita kuvataan usein niiden ei-kompositionaalisten ominaisuuksien, modifikaatorajoitusten ja muodon pysyvyyden kautta. (Evert (2008): 3). Tieteen termipankkiin kirjatun määritelmän mukaan fraseologinen yksikkö sisältää ajatuksen kollokaatiosta yhtenä kielellisen yksikön muotona, jonka piiriin mahtuvat sekä kielen ulkoisesti että sisäisesti motivoitunut kokonaisuudet (ks. määritelmä alaviitteessä 4 sivulla 9).

Merkityksen muodostuminen näyttäisi toteutuvan usealla tasolla. Heid (1999) viittaa Hausmannin yhteisesiintymistä koskeviin tutkimuksiin 1980-luvulta ja toteaa, että yhteisesiintymän osilla on toisiinsa erityinen suhde, jossa toinen elementti määrittää toista. Merkityksen muodostumisessa ei tosin ole täysin selvää, miten määrittäyssuhde suuntautuu ja mikä elementti tai sane toimii yhteisesiintymässä määrittäjänä ja mikä määritettävänä (vrt. vastavuoroisuus alaluvussa 2.2.3). Hausmannin mukaan hakusana on kuitenkin luonteeltaan autosemanttinen, toisin sanoen sillä on itsenäinen kollokaatin merkityksestä riippumaton merkityksensä, kun taas sen synsemanttinen kollokaatti on riippuvainen hakusanan merkityksestä ja sen merkitys voidaan tulkita vain suhteessa koko sanayhdistelmään (Hausmann (2004), viittaus teoksessa Seretan (2011): 13). Hakusanan merkitys ohjaa tämän teorian mukaan kokonaisuuden semanttista tulkintaa ja yhteisesiintymä on siten asymmetrinen. (Seretan (2011): 13). Tämän perusteella Hausmannin teorian mukaista yhdyssanaa ei voi pitää itsenäisenä kollokaationa, mikä on eräs käsillä olevan tutkimuksen herättämistä kysymyksistä, johon palataan työn johtopäätöksissä luvussa 8.

Yhteisesiintymä tuottaa uusia merkityksiä

Kollokaation merkitys on lausuman merkityksen erikoistumista (Sinclair et al. (2004): xxii). Täysin kompositionaalisen lausuman elementeillä voi olla yhtä suuri esiintymistajuus aineistossa kuin sellaisen lausuman elementeillä, joiden esiintyminen yhdessä vaikuttaa intuitioon nojaavan tutkijan mielestä merkityksellisemmältä. Keskustellessaan Sinclairin (JS) kanssa Teubert (WT) tarjoaa tästä esimerkkinä sanayhdistelmät *previous* ja *night* ja toisaalta *harsh criticism*. Jälkimmäisessä tapauksessa sanojen esiintyminen yhdessä rajoittaa kummankin merkitystä (Sinclair et al. (2004): xxii):

WT: But still the fact remains that even in large corpora you come across co-occurrences that you intuitively consider to be a unit of meaning (e.g. harsh criticism), and others, like previous night, which is just the addition of previous and night. Both may occur with the same frequency. But in harsh criticism we encounter a reduction of the meaning of harsh and of criticism; is this not the essential feature of a true collocation?

JS: Yes, though I would call it specialisation rather than reduction. (Sinclair et al. (2004): xxii)

Adjektiivit voivat määrittää suurta määrää substantiiveja, joten niiden merkitys joustaa substantiivin merkityksen mukaan. Substantiivin valinta aktivoi siten adjektiivin merkityskentästä osan, joka yhdessä substantiivin kanssa tuottaa idioomaattisen lausuman kyseisessä tilannekontekstissa. Tässä tutkimuksessa oletetaan,

että termikeskeisiä asiategstejä käännettäessä substantiivin valinta ohjaa sen kanssa käytettävän verbin valintaa ja käytettävissä olevien vaihtoehtojen joukko rajautuu käännettävän substantiivin oletetun merkityksen perusteella. Tuloksena oleva sanayhdistelmä kuvastaa siten substantiivin merkitystä verbin valinnan kielen käyttäjän leksikossa tai ajatusmaailmassa tai vaikkapa kääntäjän apuvälineessä. Yhteisesiintymisen mekanismi mahdollistaa lausuman merkityksen muodostumisen juuri kyseiseen kielenkäyttötilanteeseen sopivaksi ja merkityksen muodostumisen voi ennustaa tilastollisen tarkastelun ja todentaa intensionaalisen tason tarkastelun avulla.

Pasanen viittaa Niemikorven tutkimukseen, jonka mukaan substantiivit ja yhdyssanat ovat erikoiskielissä yleiskieleen verrattuna yleisempiä ja verbit harvinaisempia (Pasanen (2009): 40). Erikoiskielissä syntaktiset rakenteet N+V, N+Adj ja N+N ovat Euroopan kielissä huomattavasti taajemmin esiintyviä kuin V+Adv tai Adj+Adv (Heid (2001): 789).

Adjektiivisten vahvistuussanojen merkitys voi muuttua ajan mittaan ja niissä voi tapahtua syntaktisia muutoksia (Partington (1993)). Tutkiessaan vahvistuussanojen merkityksen diakronista muutosta Partington havaitsi kivettyneitä, leksikaalistuneita sanayhdistelmiä, jotka ovat lakanneet olemasta produktiivisia (esimerkiksi *stone dead*, *stone deaf* ja *stone cold*). (Partington (1993)) Suomen kielessä näistä ensimmäinen ja viimeinen rakentuvat englannin tavoin: *kuollut kuin kivi* ja *kylmä kuin kivi*, mutta **kuuro kuin kivi*.

Over the centuries, the items which have been described as delexicalized intensifiers have experienced a steady decline in their ability to function as separate lexical choices. Probably the clearest example of this process is the behavior of *absolutely*, which, as mentioned in section 2 above, used to mean “separately, independently (O.E.D.). However, judging from the evidence from the modern corpus, the main function of the word today seems to be to modify terms already marked for strength or superlativity, such as *enchanting*, *delighted*, *shocking* and *appalling* . . . The item *absolutely* seems to add little independent meaning to the adjectival group, apart, of course from that conveyed by intensification. It functions purely in combination with the terms it modifies. (Partington (1993): 187)

Martin (1992) määrittelee leksikaalisen kollokaation (lexical collocation) edustavan sellaista merkityksen muodostumista, jossa yhteisesiintymä on ennustettavissa maailmantiedon tai tilannekontekstin hallinnan kautta. Säädoskielen kääntäjän kannalta tämä tarkoittaa erityistä vaatimusta alan, tekstilajin ja aihepiirin hyvän hallinnan vaatimusta. Vastaava käsite terminologian puolella on konseptuaalinen kollokaatio, jossa määritely käsite avaa paikan toiselle käsitteellisen luokan edustajalle. Esimerkiksi *disease* mahdollistaa yhteisesiintyjäkseen *lekseemin*, joka kuuluu luok-

kaan [+ CAUSE] (siteeraus Heid (2001): 794–795). Generatiivisen kielentutkimuksen traditiossa vastaavasti käsitepiirrejärjestelmän syntaktiset rajoitukset säätelevät sitä, millaisissa rakenteissa leksikaalinen elementti voi esiintyä tai millaisia semanttisia piirteitä omaavia yksiköitä rakenne edellyttää. Verbi voi saman rakenneyksikön sisällä esimerkiksi vaatia subjektiksi substantiivin, johon liittyy jokin tietty luontaispiirre, vaikkapa elollisuus. Crystalin esimerkissä **the stone slept* verbillä to *sleep* on piirrematriisissaan syntaktinen rajoitus [+ANIMATE]. (Crystal (2003)). Yhteisesiintymän elementtien sanavalinnat ennustavat kontekstia ja konteksti ennustaa sanavalintaa.

Tutkimuksen tavoitteena on osaltaan arvioida, päästäänkö tutkimuksen määrällisin menetelmin tarkastelemaan merkityksen muodostumista.

Läpinäkyvyys ja kompositionaalisuus

Läpinäkyvyydellä tarkoitan tässä työssä kielellisen elementin, esimerkiksi sanan, symbolin tai välimerkin merkityksen yleistä tunnistettavuutta. Kuullessamme puhetta tai nähdessämme tekstiä, mielessämme aktivoituu jonkinlainen vastine, representaatio, tekstiyhteys, tilannekonteksti tai muu assosiaatio. Yksilön tasolla assosiaatio varmasti riippuu kielen käyttäjän kokemusmaailmasta, joista yhteisön tasolla muodostuu yhteinen, jaettu tieto elementin merkityksestä. Yhteinen tieto kirjautuu teksteinä ja tallentuu puheena materiaaliksi sanakirjoihin ja käsitesanakirjoihin. Vapaan valinnan periaatteiden (ks. luku 2.5) mukaiset yhteisesiintymät ovat kompositionaalisia. Pelkkä yhdisteleminen sinänsä ei Sinclairin mukaan välttämättä muuta kollokaation osien tavanomaista merkitystä: ”...is usually called a collocation, a frequent co-occurrence of words; it does not have a profound effect on the individual meanings of the words, but there is usually at least a slight effect on the meaning, if only to select or confirm the meaning appropriate to the collocation, which may not be the most common meaning.” (Sinclair (1996): 80). Kollokaatio on siten jollakin tapaa ylimääräistä, kerrostunutta merkitystä, joka laajentaa tapoja käyttää kyseistä kielellistä elementtiä.

Yhteisesiintyminen leksikaalisena ilmiönä on jatkumo, joka alkaa vapaan valinnan periaatteen mukaisesti valikoitujen komponenttien yhdistelmästä (esimerkiksi numeraali ja substantiivi) suljetun valinnan periaatteen mukaisesti valikoitujen elementtien yhdistelmään (esimerkiksi idiomi). On siis hyväksyttävä, että vaikka ilmiön kiinnostavuus ja merkittävyys kielen käyttäjän kannalta liittyy jatkumon eikompositionaaliseen päähän, jossa helpoiten luodaan tietoisesti erilaisia tunnetiloja (tuttuutta, epämukavuutta, iloa tai vaikkapa hämmästyttä), on ilmiö sama jatkumon toisessakin päässä. Sanoja ja muita kielellisiä elementtejä sijoitetaan kielessä peräkkäin tilannekohtaisesti määräytyvien lainalaisuuksien mukaisesti. Osittaisella

kompositionaalisuudella (L’Homme, viittaus teoksessa Heid (2001): 794) ymmärrän tässä tarkoitettavan epätarkkuutta merkitysten muodostumisessa ”sana tarkoittaa suunnilleen jotakin, ilmaus on sinne päin”. Kuten Heid toteaa, erikoiskielissä osittaista kompositionallisuutta ei juuri esiinny, koska termit on yleensä määritelty riittäväällä tarkkuudella (Heid (2001): 794).

Sag ym. (2002) käsittelevät moniosaisen ilmausten idiosynkrasiaa ja määrittelevät käsitteen seuraavasti: ”idiosyncratic interpretations that cross word boundaries (or spaces)” (Sag et al. (2002): 2). Moniosaiset elementit jakautuvat hänen työssään kahteen luokkaan: leksikaalistuneet fraasit (lexicalized phrases) ja institutionaaliset fraasit (institutionalized phrases). Sag ym. huomauttavat myös, että Jackendoffin (1997: 156, viite Sag et al. (2002): 2) arvion mukaan kielen puhujilla on sanavarastossaan moniosaisia elementtejä yhtä paljon kuin yksittäisiä sanoja. Leksikaalisten fraasien luokkaan kuuluvissa kollokaatioissa voi olla mukana elementtejä, joita esiintyy kielessä vain näiden fraasien osana (esim. *to and fro*). Institutionaaliset fraasit ovat syntaktisesti ja semanttisesti kompositionaalisia, mutta ne esiintyvät jossakin tiettyssä tilannekontekstissa siinä määrin usein, että ne konventionaalistuvat kielen puhujien yhteisössä ja ne ovat huomattavasti tavallisempia kuin saman käsitteen vaihtoehtoiset konstruktiot. Tilastollinen yhteys käsitteen ja tarkoitteen välillä voi muodostua niin voimakkaaksi, että sanayhdistelmä leksikalisoituu ja sen osien kompositionaalisuus vähenee. Esimerkkinä moniosaisesta institutionalisoituneesta fraasista on *traffic light*, jolla viitataan ainoastaan liikenteenohjauksen työkaluun, liikennevaloon, vaikka se voisi yhtä hyvin merkitä myös esimerkiksi liikennevälineen kääntymistarkoitusta tarkoittavaa signaalia, eli vilkkua, jolla on puolestaan oma vaikiintunut muotonsa *turning signal* (Sag et al. (2002): 3–8).

Tarkastelen yhdyssanoja kollokaation erikoistapauksena. Yhdyssanan osat ovat erillisinä kompositionaalisia, mutta yhdistettäessä kokonaisuuden merkitys ja käytötarkoitus muuttuvat: ”When the elements of compound words collocate, they form new lexical items: house and boat coming together in both houseboat and boathouse, each with a distinct meaning and use.” McArthur, verkkosanakirja¹⁵. Myös Sinclairin mukaan englannin yhdyssanojen merkitys on eri kuin niiden osien summa, jolloin yhdyssanan muodostavat kielelliset elementit ovat lähempänä muita rakenteita, joissa merkitysyksikkö on ortografista sanaa laajempi. Tällaisia muita rakenteita ovat idiomit (idioms), partikkeliverbit (phrasal verbs), kiinteät fraasit (fixed phrases), joustavat fraasit (variable phrases), kliseet (clichés), sananlaskut (proverbs) ja monet tekniset termit (technical terms) (Sinclair ja Carter (2004): 25). Baldwin ym. (2010) huomauttavat, että moniosaisen ilmauksen komponentit esiintyvät hyvin erilaisissa konteksteissa kuin moniosainen ilmaus kokonaisuutena (Baldwin ja Kim (2010): 24).

¹⁵Concise Oxford Companion to the English Language. 1998. Retrieved May 31, 2010 from Encyclopedia.com: <http://www.encyclopedia.com/doc/1O29-COLLOCATION.html>

Vakiintumat ovat äidinkieliselle kielen käyttäjälle läpinäkyviä niiden tulkinnan puolella, mutta niiden tuottaminen pyynnöstä ilman tilannekontekstia tai tekstiyhteyttä on usein vaikeaa (Heid (1999)). Puhuttaessa tai kirjoitettaessa vakiintumat taas useimmiten tuotetaan sujuvasti. Läpinäkyvyys toimii siten äidinkieliselläkin puhujalla sujuvasti vain toiseen suuntaan. Voidaankin kysyä, muistuttaako kääntäminen toimintana enemmän vieraan kielen tuottamista kuin äidinkielen tekstin tuottamista.

Systemi, paradigma, vaihdettavuus ja lähisynonymia

Systemi on Firthin kieliteorian keskeisiä käsitteitä. ”Systemi muodostuu niistä paradigmaattisista¹⁶ vaihtoehdoista, jotka ovat valittavissa tietyn rakenteen tietyssä paikassa.” (Shore (2012): 133). Kielen paradigmaattinen akseli¹⁷ tarjoaa kielen käyttäjälle mahdolliset vaihtoehdot, joiden joukosta hän valitsee haluamansa leksikaalisen elementin osaksi tekeillä olevaa rakennetta. Akselille sijoittuvat potentiaaliset vaihtoehdot koko lausekkeen tasolla, yksittäiset vaihtoehdot kielelliset elementit sanaston tasolla tai mitkä tahansa muut tyyliin, viestin relevanssiin ja/tai muuhun tilannekontekstiin liittyvät vaihtoehdot. Tutkin kuuden tutkimusverbin paradigmaattista vaihtelua havainnoimalla niitä hakusanan yhteisesiintyjien joukossa. Tarkastelen myös adjektiivin tai adverbien vaihtelua vakiolausekkeissa ja pyrin arvioimaan vakiolausekkeiden kiinteyttä.

Kääntäjälle syntagmaattisen ja paradigmaattisen valinnan vaihtelu tarkoittaa vaihtoehdoisten käännösvastineiden varastoa, jossa vastineilla on erilaisia käännettävän yksikön tai elementin merkityksen tarkentavia tai muokkaavia ominaisuuksia.

¹⁶Paradigma; Määritelmä: kielellisten elementtien sarja, jonka jäsenet ovat samassa kielellisessä ympäristössä vaihtoehtoisia, Selite: Paradigmoja ovat esimerkiksi foneemiparadigma eli kielessä esiintyvien äänteiden joukko ja sanan (lekseemin) taivutusmuotojen sarja, esimerkiksi: kissa : kissan : kissaa : kissalla : kissalta : kissalle jne. Termiä käytetään myös yleisemmin yksiköistä, jotka voivat korvata toisensa erilaisissa kielellisissä konteksteissa, esim.: söin omenaa / söin omenan / omenasta puuttuu pala / omena tuli syödyksi / jne. Lisätiedot: VISK2008 Paradigmalla tarkoitetaan yleensä taivutusparadigmaa eli sanan kaikkien taivutusmuotojen sarjaa (» § 53). (Lähde: Tieteen termipankki 2.03.2016: Kielitiede:paradigma.)

¹⁷Paradigmaattinen suhde; Määritelmä: Määritelmä samaan kielelliseen ympäristöön sopivien kielisysteemin yksiköiden (vaihtoehtoisuus)suhde, Selite: Kahden muodon kuuluminen samaan paradigmaan (esim. sanan kirja muotojen kirjassa ja kirjasta suhde) edustaa paradigmaattista suhdetta. Paradigmaattista on myös esim. eri lekseemien vaihtoehtoisuus tietyssä syntaktisessa asemassa, johon tietynlaiset elementit sopivat: isä / bussi / äidin tilaama paketti / se tulee illalla. Sanarakenteessa taas paradigmaattinen ympäristö on muotin muuttujaosa, jossa on paikka leksikaaliselle ainekselle: ilo- / viha- / kiukku-inen. Sanan paradigmaattista ulottuvuutta edustaa myös taivutusparadigma. Lisätiedot: Roman Jakobsonin poeettisen kielen teoreemassa käytetty termi valinnan akseli (suom. Leino 1982) tarkoittaa kielisysteemin paradigmaattista ulottuvuutta. (Lähde: Tieteen termipankki 2.03.2016: Kielitiede:paradigmaattinen suhde.)

Näin yhteisesiintyjän merkitystä tarkentava ja muokkaava tehtävä toteutuu käännöstyössä. Säädoskielen luonteeseen kuuluu, että yhteisesiintymä voi sijaita kaukana pääsanastaan.

Yhteisesiintyjien osien vaihdettavuus ilman merkityksen muuttumista liittyy kiinteästi osien välisen sidoksen voimakkuuteen. Synonymiaa ja lähisynonymiaa on tutkittu suomen kielessä myös käännöksistä koostuvasta aineistosta. Jantunen tutkimus vuodelta 2004 valottaa synonyymisten astemääritteiden *hyvin*, *kovin* ja *oikein* keskinäisiä eroja ja toisaalta tyypillisiä ja epätyypillisiä kombinaatioita (Jantunen (2004)) käännöskielessä ja alun perin suomeksi kirjoitetussa aineistossa. Säädoskielessä synonymia ei ole tavoiteltavaa eikä suotavaa tekstilajin erityispiirteiden vuoksi (ks. luku 3). Erikois- ja ammattikielten johdonmukaisuus ja säädoskielessä tulkinnan systemaattisuus heikkenevät, jos säädoskielessä on turhaa vaihtelua ja samoilla käsitteillä on useita tarkoituksia tai käyttötarkoituksia (erkoiskielistä ks. luku 4).

Semanttinen preferenssi ja semanttinen prosodia

Semanttinen preferenssi ja semanttinen prosodia ovat näkökulmia kollokaation semantiikkaan (Sinclair ja Carter (2004): 35, Partington (2004)). Molemmat kytkeytyvät tilannekontekstiin: preferenssi tuo tulkintaan sanasemantiikan ja prosodia pragmatiikan näkökulman.

Semanttinen preferenssi on lekseemin ja sen kanssa esiintyvän leksikaalisen elementin semanttisen ominaispiirteiden välinen suhde. Sen kautta kielen puhujan tai kirjoittajan käytettävissä olevat sanavalinnat halutun viestin välittämisessä kaventuvat. Jantunen viittaa Sinclairin tutkimukseen, jossa *naked eye* preferoi 'näkyvyyteen' liittyviä sanoja, ja vakiintumassa sen tavallisia kollokaatteja ovat verbit *see* ja *view* ja adjektiivit *visible* ja *invisible* (Jantunen (2004): 24). Semanttinen preferenssi ulottaa merkityksen sanan ulkopuolelle sen kanssa yhteisesiintyville tai myötäesiintyville leksikaalisille tai syntaktisille elementeille.

Semanttinen prosodia kytkeytyy puhunnoksen asenteisiin (Sinclair (1996) ja niillä on siten myös diskursiivinen funktio (Stubbs (2001): 25). Semanttinen prosodia havaitaan hakusanan kanssa usein esiintyvien yhteisesiintyjien semanttisista piirteistä ja tekstiyhteydestä tyypillisesti havaituista muista sanoista. Tavallisin Sinclairin raportoima prosodiaan liittyvä tutkimustulos on negatiivisuus (Jantunen (2004): 25). Sinclair rakentaa jatkumon kollokaatiosta kolligaation kautta semanttiseen preferenssiin ja lopulta semanttiseen prosodiaan, jolloin liikutaan kielellisten elementtien kombinatoriikan problematiikasta syntaktisten kategorioiden huomioimiseen ja edelleen ensin semanttiseen ja lopulta pragmaattiseen motivointiin Sinclair (1996), kun taas Partington näkee semanttisen prosodian liittyvän konnotaatioihin (Partington (1998)) ja esittää, että semanttinen prosodia voidaan nähdä semanttisen preferenssin

alakategoriana (Partington (2004): 149). Oman näkemykseni mukaan kollokaation ja kolligaation tarkastelu perustuu toistuvien rakenteiden havaitsemiseen, kun taas semanttisen preferenssin ja prosodian puolella tarvitaan tutkijan tulkintaa esimerkiksi siitä, mikä kielessä tuottaa negatiivisen luennon.

Semanttista preferenssiä ja prosodiaa pidetään kriteereinä synonyymisten vaihtoehtojen valinnassa paradigmaattisella akselilla (Jantunen (2004), Partington (2004)). Jantunen (2004) toteaa synonyymisten astemääritteiden *oikein*, *kovin* ja *hyvin* valintaan liittyvää vaihtelua tarkastelevassa tutkimuksessaan semanttisen preferenssin ja prosodian selittävän synonyymien valintoja. Myöhemmissä oppijankieleen liittyvissä tutkimuksissaan Jantunen havaitsi, että kyseisellä astemääritteellä *oikein* on vahvasti positiivinen semanttinen prosodia: esimerkiksi suomea äidinkielenään puhuvien aineistossa 68 prosentissa havainnoista lemma *oikein* löytyi positiiviseksi luokitellusta tekstiympäristöstä, kun 28 prosenttia havainnoista oli negatiivisesta ympäristöstä. Havainnot oppijanaineistosta osoittavat semanttisen preferenssin lekseimin *oikein* käytössä ihmisiin liittyvissä luonnehdinnoissa tai kuvailussa, ja vastaavasti äidinkielisten puhujien tuottamissa teksteissä niiden osuus vastaavassa kontekstissa oli huomattavasti pienempi. Jantunen toteaa, että semanttinen preferenssi ja prosodia ovat kollokaatioita abstraktimpia ilmiöitä ja niiden huomioiminen kielenopetuksessa on siten haastavampaa. (Jantunen (2016))

Tietoinen epätyypillisen semanttinen preferenssi tai prosodia liittyy yleensä huumorin tai yllätyksellisyyden tavoitteluun. Aina ei voikaan tietää, onko epätyypillinen sanavalinta tahallista vai ei (luovuudesta käännosten sanavalinnoissa ks. Kenny (2001)). Asenteet voivat piillä tekstissä sanavalintojen sisällä vähemmän ilmipantuina (käännöstutkimuksen osalta, ks. Kemppanen (2008) ja Probirskaja (2009)). Erikois- ja ammattikielissä luovuus ei ole tarkoituksenmukaista, eikä semanttinen prosodia odotuksenmukaista. On kuitenkin mahdollista, että negatiivinen tai positiivinen prosodia liittyy joidenkin lähisyntymisten tutkimusverbien valintaan kollokaateiksi tai on seurausta valinnoista. Tarkastelen siitä syystä havaintoja myös prosodian näkökulmasta.

Hunston esittää, että semanttinen prosodia on merkityksen tavoin kontekstisidonnaista, eikä ole sanottua, että tarkastelun kohteena olevalla sanalla on samanlainen prosodia muualla kuin tutkimuskohteena olevassa kontekstissa. Esimerkiksi *positive influence* on kaksi kertaa tavallisempi hänen aineistossaan kuin *negative influence*, mutta *under the influence* kollokoi usein negatiivisesti värityneiden sanojen kanssa. (Hunston (2007): 252–254). Semanttisen prosodian havainnot eivät siten välttämättä ole yleistettävissä yleiskielestä erikoiskieliin tai päinvastoin, eivätkä erikoiskielen sisällä eri tekstilajien välillä (Kübler ja Volanschi (2012)). Kübler ja Volanschi esimerkiksi havaitsivat, että yleiskielen negatiivinen prosodia tasoittuu erikoiskielessä. He tarkastelivat verbiä *to cause* englanninkielisissä ja ranskankielisissä

tieteellisissä teksteissä ja vahvistivat Hunstonin (2007) tuloksen verbin semanttisesta negatiivisesta prosodiasta vain kun sen avulla kuvataan ihmisiä kohtaavia tapahtumia, kun taas yleiskielessä verbillä on aikaisemmissa tutkimuksissa havaittu vahva negatiivinen prosodia (esim. Stubbs (1995): 247–250). Toinen Küblerin ja Volanschin oman tutkimukseni kannalta merkittävä tulos on, että he havaitsivat kääntäjien välttävän sellaisten verbien valintaa vastaineissaan, joilla on vahva semanttinen prosodia yleiskielessä. (Kübler ja Volanschi (2012))

Semanttisen preferenssin tarkastelu mahdollisena yhteisesiintyjän valintaan vaikuttavana tekijänä pohjautuu omassa tutkimuksessani aihealueiden kartoitukseen. Merkitykseen liittyvä tarkastelu asettuu kuitenkin lähelle valintoihin johtaneiden syiden tutkimusta. Se ei omassa tutkimuksessani kuitenkaan ole tutkimuskysymysten joukossa. Havaintoja kartoittamalla pyrin herättelemään hypoteeseja jatkotutkimusta varten ja tarkastelemaan tutkimusmenetelmän rajoituksia.

Tarkastelen merkitykseen liittyviä ominaispiirteitä siinä määrin, kun se on mahdollista, ottaen huomioon määrällisen tutkimuksen tulosten suuren määrän ja niiden laskemisessa vaadittavan manuaalisen työn. Semanttinen analyysi on myös aina tutkijan oman intuition varassa, ellei käytettävissä ole kattavaa WordNet-tietokantaa tai sentimenttianalyysiä juuri kyseisistä lekseemeistä. Kääntäjän kannalta tieto semanttisesta preferenssistä tai prosodiasta olisi kuitenkin hyödyllistä, joskin erikoiskielten kanssa työskentelevälle kääntäjälle sen tulisi olla saatavilla kääntäjän apuvälineissä, termipankeissa tai käännösmuistiin sijoitetuissa aineistoissa ja tiedon tulisi perustua erikoisalan sisäiseen ymmärrykseen. Kuten Stewart (2009) huomauttaa, korpusaineiston tulkinta on aina subjektiivista.

2.2.4 Mielleyhtymä/assosiaatio

Kollokaation ominaispiirteistä vaikeimmin tutkittavia tämän tutkimuksen menetelmien ovat mielleyhtymään ja assosiaatioon liittyvät piirteet. Kartoitan tässä niistä joitakin lyhyesti ja arvioin tutkimuksen kuluessa, miten pitkälle niitä on mahdollista havaita tutkimusaineistossa ja verrannaisaineistoissa. Rajanveto merkitykseen liittyvien ominaispiirteiden kanssa on tässä vaikeaa, ja edellä lyhyesti kuvatut semanttinen preferenssi ja prosodia voidaan yhtä hyvin luokitella mielleyhtymän otsikon alle. Ajattelen, että mielleyhtymä ja assosiaatio ovat psykolingvistiikan alan käsitteitä ja niiden kuvaamiseen tarvitaan vastaavasti psykolingvistisiä menetelmiä, jotka eivät ole tämän tutkimuksen ulottuvilla. Semanttista preferenssiä ja prosodiaa on kuitenkin pystytty kartoittamaan korpustutkimuksen menetelmin, eikä se ole poissuljettua myöskään mieleen liittyvien prosessien kuvaamisessa, vaikkakaan tässä tutkimuksessa siihen ei ole tarvittavia lähtötietoja eikä mahdollisuuksia. Tarkastelen tässä alaluvussa myös leksikaalista virittämistä (lexical priming, oma käännökseni),

aihealueen vaikutusta valintoihin ja vastavuoroisuutta.

Assosiaatio tarkoittaa yleiskielessä miellelyhtymää: kun kuulet sanan 'vesi', ajattelet ehkä janoisena juomista ja kuumana kesäpäivänä uimista – ja jos sanon 'juoda', saatat janoisena ajatella mitä tahansa nestettä, ehkä lempijuomaasi. Sanasta 'uida' tulee kaupunkilaislapselle ehkä ensimmäisenä mieleen uimahalli ja muille meri tai järvi. Assosiaatio on puhujan kokemusmaailmaan liittyvä käsite ja sitä tutkitaan psykolingvistiikan keinoin, muun muassa erilaisilla sanavalintatesteillä, silmänliike-tutkimuksella tai muilla reaktiotesteillä. ”Ihmisillä on käytössään monimuotoisia ja yksilöllisiä tapoja jäsentää semanttisen muistin sisältämää assosiativista tietoa” ja ”...semanttisten assosiaatioiden on todettu voivan helpottaa suoritusta nimeämistehävässä, mutta häiritsevän suoritusta lauseen tuottotehtävässä.” (Aro (2012): 8–14)

Kielen ymmärtämiseen liittyviä ilmiöitä on psyko- ja neurolingvistiikassa tutkittu lääketieteellisen aivokuvantamisen keinoin muun muassa muistisairauksiin liittyvien ilmiöiden ymmärtämiseksi. Kollokaatiotutkimuksessakin on pyritty hahmottamaan kollokaation syntymekanismia, koska arvellaan, että käyttäessään kieltä sen puhujat tottuvat kuulemaan ja käyttämään sanoja yhdessä toisten sanojen kanssa ja sanojen taipumus esiintyä yhdessä tekee kielellisestä viestinnästä paremmin ennakoitavaa. Ilmiötä tarkastellaan leksikon rakentumisen mekanismina, mihin liittyen tarkastelen seuraavassa Michael Hoeyn ajatuksia leksikaalisesta virittämisestä (lexical priming, ks. Hoey (2005))

Lexical priming – leksikaalinen virittäminen

Priming on terminä vielä suomen kielessä vakiintumaton. Avoimen yliopiston Psykologian englanti-suomi-sanastossa¹⁸ termille on annettu vastine **virittäminen** (tiedon prosessoinnissa), Tieteen termipankkiin kirjattu vastine sosiaalipsykologian puolella on **pohjustus** ja EU:n IATE-termipankin vastineiden joukosta löytyy viimeksi mainitun lisäksi **esiaktivointi** aihealueelta koulutus ja viestintä. Valitsen tässä työssä termin suomenkieliseksi vastineeksi vaihtoehdon **leksikaalinen virittäminen**, mutta vain välttääkseni englanninkielisen termin käytön. Nähtäväksi jää, mikä vastine lopulta vakiintuu suomalaisen tiedeyhteisön käyttöön. Teorian tavoitteena on selittää kollokaation syntymekanismia, mutta psykolingvistiikan tai neurolingvistiikan tutkimuskentillä ei vielä ole selkeää tieteellistä näyttöä ihmisen kielellisen kompetenssin muodostumisesta ja kollokaation roolista kielen tuottamisessa tai omaksumisessa aivojen toiminnan näkökulmasta.

Leksikaalisen virittämisen teoria pyrkii selittämään, miksi kollokaatio on olemassa ilmiönä ja kollokaatiot sen edustajina (Hoey (2005)) ja miten kielen puhujat liittä-

¹⁸<https://www.avoin.helsinki.fi/oppimateriaalit/psykologia/avoinsanasto.htm>

vät kielellisiin elementteihin tietoa niiden käyttöyhteydestä ja pystyvät näin aktivoimaan tiedon helpommin kielen tuottotilanteissa. Näin kohdatessaan kieltä kontekstissaan kielen puhuja alkaa käyttää itse samanlaisia rakenteita tai yhteisesiintymiä kieltä tuottaessaan. Kielen käyttäjä pyrkii toimimaan taloudellisesti ja kykenee näin käsittelemään kontekstuaalista tietoa tehokkaasti. Kollokaatioiden havaitsemisen nopeutta on mitattu leksikaalisen virittämisen kontekstissa muun muassa silmänliiketutkimuksilla (Vilkaitė (2016), Vilkaitė ja Schmitt (2017)), joissa on osoitettu sekä äidinkielisten puhujien että toisen kielen oppijoiden tunnistavan nopeammin kollokaatioita kuin satunnaisesti valittuja sanoja. Kielenoppimisen tutkimuksen lisäksi leksikaalisen virittämisen teoriaa on hyödynnetty muun muassa rekisterin tutkimuksessa selittämään rekisterille herkistymisen ilmiötä (Sardinha (2017)).

Aineistossa usein yhdessä esiintyvät elementit yhteisesiintyvät siis myös sen vuoksi, että ne kuuluvat samaan aihepiiriin. On siksi luonnollista, että kalastuslainsäädäntöä käsittelevästä tekstistä löytyy odotuksenmukaisia sanoja, kuten *kala*, *kalalaji*, *trooli*, *kalastusalue* tai *kalastusalue* ja että niiden muodostamat yhteisesiintymät osoittautuvat tilastollisesti merkitseviksi. Aiheperusteisia yhteisesiintymiä käytetään muun muassa tekstin luokittelun apuvälineinä. Ne nousevat myös esiin etsittäessä yhteisesiintymiä jäsennetystä aineistosta syntaktisen suodattimen avulla, jossa hakukriteereissä annetaan rajoituksia sallituista sanaluokkayhdistelmistä esimerkiksi rajoittamalla sallittuja sanaluokkia tai komponenttien lukumääriä. (Manning ja Schütze (1999))

Leksikaalinen virittäminen voi selittää mekanismia, jolla kääntäjä valitsee kohdekielisen vastineen käännettävässä tekstissä esiintyvälle sanalle tai termille. Voidaan ajatella, että käännettäessä mekanismi ei toteudu samalla tavoin kuin suoraan äidinkieltä kirjoitettaessa. Leksikaalinen virittäminen voisi siten toimia interferenssiä selittävänä tekijänä. Leksikaalisen virittämisen tutkimukseen käänösprosessissa tarvitaan kuitenkin käänöstehtävän tilannekontekstin parempaa kartoitusta sekä esimerkiksi silmänliiketutkimusta tai kääntäjän työn tekemistä näkyväksi psykolingvistiikan menetelmien avulla.

Oman tutkimukseni näkökulmasta leksikaalisen virittämisen tai leksikaalisen virittäytymisen teoria on oleellinen yhteisesiintymien vakiintumisen ja niiden syiden selittämisessä. Erityisen mielenkiintoista on sen implikoima yhteisesiintymien korvaaminen uusilla sen seurauksena, että kääntäjä joutuu alttiiksi uusille tavoille ilmaista asia. Tämä olisi vaihtoehtoinen selitys epätyypillisten vastineiden vakiintumisen hypoteesille, jonka mukaan yhteisössä jo valitut vastineet jäävät ennakkotapauksina voimaan, kun kyseessä on säädösteksti. Sen sijaan leksikaalisen virittämisen teorian perusteella voidaan esittää, että jouduttuaan alttiiksi tietyille yhteisesiintymille kääntäjä käyttää niitä myös itse kääntäessään tiedostamattomasti.

Sama aihepiiri

Auli Hakulinen ja Jussi Ojanen määrittivät Firthiin viitaten vuonna 1970 ilmestyneessä teoksessaan ”Kielitieteen ja fonetiikan terminologiaa” (Hakulinen ja Ojanen (1970)) vakiintumaksi kutsumansa ilmiön seuraavasti: Vakiintuma on a) ’kieliopillisen elementin kanssa syntagmaattisessa konstruktiossa esiintyvä muoto’, esimerkiksi *juoda* + substantiivit jotka ovat nestemäisessä muodossa, b) ’sanan syntagmaattisten assosiaatioiden joukko’, ’todennäköisyys, jolla sana a tai b esiintyy jonkin muun sanan välittömässä yhteydessä, esimerkiksi *haukkua – koira, pimeä – yö* (Hakulinen ja Ojanen (1970): 118). Tieteen termipankissa myötäesiintymäsuhde¹⁹ määritellään samalla tavoin. Tämän tutkimuksen käsitteistössä määritelmä kuvastaa vakiintuman joustavuutta: koska sen ytimenä on merkitys, ei sitä voida kiinnittää kielellisen elementin muotoon. Käsitteiden siten myös verbikantaiset substantiivit ja muut johdannaiset (*tanssia-tanssiminen-tanssiva*) saman merkityksen erilaisiksi ilmenemismuodoiksi määritelmän mukaisesti.

Ensimmäisen määritelmän mukaan esimerkiksi verbin (kuten Hakulisen ja Ojanen esimerkissä) merkitys määrää sen kanssa tarvittavien elementtien (esimerkissä substantiivin) valintaa. Sanoilla nähdään olevan sisäänrakennettuja ominaisuuksia, kuten verbillä *juoda*, jotka automaattisesti rajaavat mahdollisia yhteisesiintymiä, kuten esimerkissä juomisen kohde voi olla olomuodoltaan ainoastaan nestemäinen. Kun verbi aktivoituu, aktivoituu myös sen semanttinen rooli (Aro (2012): 9). Toisaalta myös puheenaihe määrittelee viestinnässä tarvittavia yksiköitä, koska myös tilastollisesti saman aihealueen sanasto esiintyy usein samassa puhe- tai tekstiyhteydessä. Määritelmä kuitenkin poikkeaa Firthin näkemyksestä, jonka mukaan ”Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words.” (Firth 1957: 196, siteerattu Xiao (2015): 107). Käsitte pohjaisia (conceptual) yhteisesiintymiä ei löydy puhtaasti tilastollisten menetelmien avulla (Xiao (2015): 107). Hakulisen ja Ojanen toinen määritelmä rinnastaa assosiativisen sidoksen tilastolliseen todennäköisyyteen (ks. tämän luvun ensimmäinen alaluku 2.2.1) ja osoittaa, että kollokaation käsite kattaa jatkumon tilastollisesta näkökulmasta puhtaaseen assosiativisuuteen, katso myös osio vastavuoroisuus alla.

Varsinaisen tutkimusaineiston (EU) ja ensimmäisen verrannaisaineiston (SL) (esitely tarkemmin luvussa 4), aihepiirit ja korpusten sisältämien asetusten ja lakien aihealueet on dokumentoitu yleisellä tasolla, mutta aineistojen analyysiin tietoa ei aikanaan ole valitettavasti sisällytetty. Tieto olisi tuonut mahdollisuuden arvioida ai-

¹⁹Myötäesiintymäsuhde; Määritelmä: Kahden sellaisen kieliopillisen rakenteen välinen (moleminpuolinen) suhde, joissa on samat peruselementit ja jotka ovat muunnettavissa toisikseen transformaation avulla, esim. tyttö tanssii tanssiva tyttö tytön tanssiminen, engl. co-occurrence relation (Lähde: Tieteen termipankki 2.03.2016: Kielitiede:myötäesiintymäsuhde.)

healueen vaikutusta kollokaattien valikoitumiseen, mutta analyysi on tehty manuaalisesti. Analyysin tarkentaminen näiltä osin tarjoaisi lisätietoa merkityksen muodostumisesta, joka on tämän tutkimuksen ulottuvilla vain tutkimusikkunan sallimassa laajuudessa.

Vastavuoroisuus

Tutkimukseni kannalta vastavuoroisuus tarkoittaa kysymystä siitä, kumpi tai mikä kielen komponentti tai kielellinen elementti on lähtökohta yhteisesiintyjien tai kollokaattien valikoitumisessa ja missä määrin valinta voisi tapahtua yhtäaikaaisesti. Kääntäjän käytettävissä olevien termipankkien kehittyessä hakuja voidaan tehdä useammalla kuin yhdellä hakusanalla ja tavoitteena on, että fraasit, lausekkeet, lauseet tai virkkeet tuottuvat laajempina vastineyksikköinä. Apuvälineiden ja hakukoneiden kehittämisellä on erityisen suuri merkitys erikoiskielten kääntäjälle työkalujen sisällön kasvaessa ja vaihtoehtojen karttuessa.

Mel'čuk (1998) määrittelee kollokaation muodostuvaksi perusosasta (base) ja sen kollokaatista (collocate). Perusosa on semanttisesti määräytyvä ja sen kollokaatti leksikaalisesti valikoituva. Kollokaatitieto (combinatorial information) tarjotaan sanakirjoissa perusosan yhteydessä, joka on tavallisesti substantiivi: substantiivin semantiikka ohjaa koko lausekkeen semantiikkaa. Esimerkiksi substantiivi *rain* saa kollokaateikseen määreet *heavy* ja *cats and dogs*, joista kumpikaan ei kuitenkaan ole substantiivin *wind* kollokaatti – *strong* sen sijaan on. (Mel'čuk (1998): 23–53) Mel'čuk näkee siis kollokaation seuraussuhteena, jossa yhden elementin valinta johtaa toisen elementin valintaan. Pohtiessaan tilastollisten menetelmien kehitystä ja suuntaa myös Gries kiinnittää huomiota mahdollisuuteen, että kollokaatit olisivat symmetrisyyden sijaan yksisuuntaisia (directional), mitä useimmat tilastolliset menetelmät eivät oleta (Gries (2013): 141). Kahden leksikaalisen elementin välinen sidos voidaan Yuretin (1998) mukaan määrittää kahdenlaisen tiedon varassa: kieliopillisten rajoitusten (tiedämme, että adjektiivit määrittävät substantiiveja) ja toisaalta sanavalintoihin liittyvien rajoitusten (tiedämme, että verbi *juoda* vaatii nestemäisen objektin). Yuretin käsite **leksikaalinen vetovoima** (lexical attraction) liittyy näistä jälkimmäiseen rajoitukseen, jonka voimakkuutta mitataan tilastollisin menetelmin. (Yuret (1998))

Ei ole kuitenkaan itsestään selvää, että esimerkiksi kollokaation pääasanaksi mielletty substantiivi on kollokaation muodostumisen lähtökohta. Bartschin mukaan verbi voi myös määrittää substantiivin merkitystä (Bartsch (2009)). Näkemys on aineistoni valossa mielenkiintoinen, sillä tutkittavien verbien joukossa on sekä lähisynonyymisiä (*suorittaa, tehdä, toteuttaa*) että perusmerkitykseltään muista eroavia verbejä (*soveltaa, säättää*). Näiden verbien valinta tutkittavan substantiivien yhteyteen

saattaa viestittää kyseisen substantiivin merkityskentän muuttumisesta.

Eräs mahdollinen tarkastelukulma kollokaatin pääsanan ja kollokaattien väliseen suhteeseen on niiden frekvenssi kielessä yleensä tai tutkimuksen kohteena olevassa populaatiossa. Sinclair (Sinclair (1991), Sinclair et al. (2004)) esittää, että kompositionaalisessa kollokaatiossa kollokaatti eli yhteisesiintyjä on aineistossa tavallisempi kuin hakusana. Käsittelen näkökulman assosiaation yläotsikon alla, vaikka sen käsittely ja lähtökohta liittyvät kiinteästi määrälliseen tarkasteluun. Sinclairin termi tällaiselle kollokaatiolle on **upward collocation**, josta käytän nimitystä **yleisempään elementtiin suuntautuva yhteisesiintymä** (YLSU). Tällainen kollokaatio syntyy sen taaajan käytön ansiosta. YLSU muistuttaa **satunnaista yhteisesiintymää** (Firthin termi **common collocation**, Sinclairin vastaavasti **casual collocation**), jossa sanayhdistelmän osat valikoituvat aihealueen perusteella. Se näyttäisi lähestyvän vapaan valinnan periaatetta (ks. luku 2.6), jossa kollokaatio ei määräydy merkityksen vaan tavan perusteella. Kompositionaalisessa kollokaatiossa kollokaatti vaikuttaa pääsanansa merkitykseen vahvistavasti, muttei rajaa tai tarkenna sanakirjamerkitystä. Sinclairin esimerkki edellä mainitun adjektiivin *confused* tavallisesta kollokaatista on ad-adjektiivi *very*, joka on englannin kielessä tavallisempi kuin *confused* ja voi vahvistaa lähes mitä tahansa suhteellista adjektiivia. Kollokaatio *very confused* on esimerkki yhteisesiintymien jatkumolla sijaitsevasta kompositionaalisesta yhteisesiintymästä, jonka elementit yhdistetään vapaasti vapaan valinnan periaatteen mukaisesti (ks. luku 2.6). Tällaiset yhteisesiintyjät ovat Sinclairin mukaan usein kieliopillisia elementtejä tai yläkäsitteitä (Clear (1993): 281).

Harvinaisempaan kielelliseen elementtiin suuntautuva yhteisesiintymä (HARSU) on vastineeni Sinclairin käsitteelle **downward collocation**. HARSU ovat Sinclairin ajatusmallin mukaan ja omaa termistöäni soveltaen **myötäesiintymiä**, joiden yksittäisten komponenttien merkityksiä ei voi tulkita jakamalla sanayhdistelmän osiinsa ja tulkitsemalla jokainen osa perusmerkityksessään. Kollokaatio on ei-kompositionaalinen. Niissä kollokaatioissa yhteisesiintyjä on aineistossa harvinaisempi kuin hakusana. Ne ovat myös kielessä sillä tavoin **merkittäviä** (Sinclairin termi **regular collocation**), että ne on opittava kokonaisuuksina. Selkein esimerkki HARSUsta on idiomi.

Intensionaalisen kollokaation komponenttien esiintymisellä yhdessä on merkitystä muodostava rooli. Silloin pelkästään se seikka, että elementit esiintyvät yhdessä tuo ilmaisuun merkitysvivahteen, joka muodostuu hakusanan merkityksen rajaamisesta kollokaation avulla (syntyy koko lausuman merkitysvivahde eli semanttinen prosodia). Voidaan myös luoda kokonaan elementtien tai osien sanakirjamerkityksistä riippumaton uusi merkitys (esim. idiomaattiset ilmaukset). (Sinclair et al. (2004)) Sinclairin *confused*-esimerkkiä (yllä) jatkaen muodostamalla vakiintuma adjektiivista *confused* ja sen kanssa esiintyvistä ad-adjektiiveista lähestytään sanayhdistelmien

muodostaman jatkumon ei-kompositionaalista päätä. Jos adjektiivia määrittää *utterly*, vaikuttaa se adjektiivin *confused* merkitykseen ja tuloksena vakiintuman *utterly confused* semanttinen prosodia on aina negatiivinen (Sinclair et al. (2004): xxiii); astemääritteiden tutkimuksesta katso myös Jantunen (2004). Merkitsevä (significant) vakiintuma on semanttinen kategoria, jossa merkitys muodostuu vakiintuman osien (enemmän tai vähemmän kompositionaalisten) merkitysten summasta, mutta on kuitenkin kielen käyttäjälle tuttu eli toteutuu kielessä riittävän usein.

YLSU = yleisempään elementtiin suuntautuva yhteisesiintymä, jossa hakusana/noodi on aineistossa harvinaisempi kuin yhteisesiintyjä (yhteisesiintyjä on aineistossa tavallisempi kuin hakusana). On kompositionaalinen ja motivoituu merkityksen perusteella, on tavallinen, syntyy tilastollisista syistä. Tässä tutkimuksessa yhteisesiintymä.

Jos yhteisesiintyjän kriteerinä pidetään yhteisesiintyjän frekvenssiä, riippumatta siitä, onko se tavallisempi tai harvinaisempi kuin hakusana (noodi), olisi aineistossa potentiaalisia YLSUja kaikki yhteisesiintymät, joiden komponenttina on jokin lakikielessä taajaan esiintyvä verbi, kuten *säättää*. Tarkastelen kysymystä tarkemmin tutkimusosiossa luvussa 6.

kun taas

HARSU = harvinaisempaan elementtiin suuntautuva yhteisesiintymä, jossa hakusana/noodi on aineistossa tavallisempi kuin yhteisesiintyjä (yhteisesiintyjä on aineistossa harvinaisempi kuin hakusana). On merkittävä, syntyy merkitykseen liittyvistä syistä. Tässä tutkimuksessa vakiintuma.

Merkittävä kollokaatio (esim. idiomi) voi olla kielessä harvinaisen, mutta sen osat voivat esiintyä aineistossa useammin yhdessä kuin erikseen. Satunnainen kollokaatio voi taas olla kielessä hyvin tavallinen ja sen osat voivat esiintyä aineistossa useammin yhdessä kuin erikseen. Molemmat täyttävät ehdon kollokaation riittävästä esiintymistäajuudesta eli voivat olla merkitseviä yhteisesiintymiä, mutta vain merkittävä kollokaatio muodostaa osistaan erityisen merkitystä kantavan yksikön, vakiintuman. (Sinclair et al. (2004): xxii–xxiii)

Aineistosta ei tekstilajin luonteesta johtuen ole odotuksenmukaista löytää ei-kompositionaalisia HARSUja, vaikka niitä ei säädöskielessäkään voi pitää täysin poissuljettuina. Oikeuskielessä saattaa tosin olla historiallisia vakiofraaseja, joiden merkitystä ei enää komponenttien perusteella pysty tulkitsemaan. Niiden löytäminen tämän tutkimuksen aineistoista on kuitenkin epätodennäköistä. Tutkimusosiossa tarkastelen kuitenkin kysymystä siitä, kumpi yhteisesiintymän komponenteista esiintyy korpuksessa yhteisesiintymän ulkopuolella toista taajemmin.

Tarkastelen Sinclairin merkitseviä ja harvinaisia yhteisesiintymiä merkityksen muodostumisen kannalta, koska ne näyttäisivät puoltavan ajatusta kahdenlaisista

kollokaatioista: merkityksen perusteella muodostuvista myötäesiintymistä ja tilastollisista syistä muodostuvista yhteisesiintymistä. Ilman ei-komponentiaalisuuden vaatimusta potentiaalisia HARSUja olisivat muuttumattomat vakiolausekkeet, esimerkiksi Suomen säädöskokoelmassa lakitekstin aloittava vakiolause *Eduskunnan päätöksen mukaisesti säädetään*:. Sinclairin esimerkit ovat adjektiivien ja substantiivien yhdistelmiä ja teoria soveltuu siten paremmin substantiivivia määrittävän lauseenjäsenen tarkasteluun. Tämän tutkimuksen tutkimuskohteena olevat kollokaatit ovat verbejä, joten teorian sovellettavuus on kyseenalaista. Kuitenkin ajatus kollokaation hakusanan ja kollokaatin yleisyyden merkityksestä kollokaation muodostumiseen on kiinnostava, ja tarkastelen siksi kysymystä tutkimusosiossa luvussa 7 kartoittamalla tarpeelliset tunnusluvut ja sivuan siinä yhteydessä tarkemmin tutkimusverbien ja tutkimuskohteen muodostamien kollokaatioiden sijoittumista HARSUjen (downward collocation) tai YLSUjen (upward collocation) joukkoon tutkimusaineistossa ja ensimmäisessä verrannaisaineistossa.

2.3 Konventionaalisuus ja norminmukaisuus

Konventionaalisuudella tarkoitan yhteisön sisällä syntyvää tai olevaa yhteisymmärrystä jostakin asiasta. Tieteiden toimijoiden yhteisymmärrys käsitteistä ja termien tarkoitteista kirjataan Tieteen termipankkiin²⁰ ja kollektiivinen tiedon dokumentointi korvaa ihmisten välisen sopimuksen: asiasta on sovittu yhdessä. Tieteen termipankin määritelmän mukaan 'konventio' on kielitieteessä **vakiintunut käytänne**²¹ ja 'konventionaalinen' tarkoittaa **sopimuksenvaraista**²². Erikoiskielten ja ammattikielten konventiot muovautuvat yhteisön toiminnan perusteella ja syntyy hiljaista tietoa siitä, millaista alan kieli on. Erikoiskielten kannalta konventionaalisuus tuottaa fraseologista tarkentumista ja yhteisesiintymisen kannalta niiden vakiintumista. Käännöstehtävässä **tekstikonvention** ja **normien** noudattaminen tuottaa tavallisesti **hyväksyttävän**²³ käännöksen.

Vakiintuman tehtävä viestin tuottamisen tai ymmärtämisen prosessissa on a) rajata se semanttinen kenttä tai aihealue, jolla viesti liikkuu ja b) viestittää sanojen yhdistelemisen periaatteiden noudattamisen tai noudattamatta jättämisen kautta vastaanottajalle tieto siitä, onko viesti konventionaalinen eli periaatteita noudattava vai onko se epäkonventionaalinen. Konventionaaliset, odotustenmukaiset viestit ovat

²⁰<http://tieteentermipankki.fi>

²¹Tieteen termipankki 4.06.2018: Kielitiede:konventio.

(Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Kielitiede:konventio>.)

²²Tieteen termipankki 4.06.2018: Kielitiede:konventionaalinen.

(Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Kielitiede:konventionaalinen>.)

²³Tieteen termipankki 4.06.2018: Kielitiede:hyväksyttävyyys.

(Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Kielitiede:hyväksyttävyyys>.)

yllätyksettömiä, helppoja ja nopeita prosessoida, kun taas ei-konventionaaliset viestit vaativat vastaanottajalta asiayhteyden havaitsemista ja yhdistämistaitoa (Stubbs (1996)). Normit syntyvät yhteisön päätöksellä, ne ovat sitovampia kuin konventiot kirjataan termipankkeihin ja alan sanakirjoihin tai sanastoihin (säädöskielen erityispiirteistä katso luku 3). Kääntäjän käyttämät apuvälineet toimivat tietynlaisina normeina käännöstehtävän suorittamisessa, koska voidaan olettaa, että käännösmuistiin tai konekääntimeen on syötetty jo hyväksytyjä vastineita ja käännösvaihtoehtoja. Tarkastelen tutkimuksessani kysymystä EU:n IATE-termipankin sisältämiä tutkimuskohteen vastineita kieliparissa englanti-suomi saadakseni käsityksen siitä, miten laaja variaatio termin hyväksyttävillä vastineilla EU:n käännöstoimessa on nykyisin.

2.4 Tekstiyhteys ja tilannekonteksti

Päivi Pasanen kuvaa terminpoinintaa käsittelevässä väitöskirjassaan (Pasanen (2009)) kontekstin kolme tasoa: maailmantiedon tason, erikoisalan tason ja lingvistisen tason. Maailmantieto on kielen käyttötilanteessa pysyvä, staattinen konteksti, erikoisalakonteksti puolestaan on muuttuva, dynaaminen taso ja lingvistinen taso sisältää tekstin tai korpuksen termiä ympäröivän **tekstikontekstin**. Maailmantieto ja erikoisalatieto ovat termintunnistuksessa välttämättömiä lingvistisen kontekstin lisäksi. (Pasanen (2009): 65–66). Samoin kääntäjä tarvitsee työssään kaikkia kolmea tasoa. Tekstimassasta tarkasteltavan elementin välittömässä läheisyydessä (sanojen tekstuaalista lähiympäristöä) käytetään myös termiä **koteksti** (käännöstieteellisen tutkimuksen osalta esimerkiksi Jantunen 2004 (Jantunen (2004): 13 ja Sinclair (1991)) tai suomenkielistä termiä **tekstiyhteys**²⁴.

Käytän tässä tutkimuksessa säädöstekstin tuottamisen tai tulkinnan tilanteeseen ja olosuhteisiin liittyvistä asioista termiä **tilannekonteksti** (context of situation) viittaamaan niihin tekijöihin, joiden avulla voidaan kuvata kielen käytön tilanteen keskeiset elementit, mukaan lukien itse lausuma:

My view was, and still is, that ‘context of situation’ is best used as a suitable schematic construct to apply to language events, and that it is a group of related categories at a different level from grammatical categories but rather of the same abstract nature. A context of situation for linguistic work brings into relation the following categories:

A. The Relevant features of participants : persons, personalities.

(i) The Verbal Action of the participants

²⁴Tekstiyhteys; Määritelmä: tarkasteltavana olevan kielenaineuksen välitön kielellinen ympäristö. Lähde: Tieteen Termipankki 18.5.2017: Kielitiede:tekstiyhteys. (Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Kielitiede:tekstiyhteys>.)

(ii) The Non-Verbal Action of the participants

B. The relevant objects

C. The Effect of the Verbal Action.

Contexts of situation and types of language function can then be grouped and classified. A very rough parallel to this sort of context can be found in language manuals providing the learner with a picture of a railway station and the operative words for travelling by train. It is very rough. But it is parallel with the grammatical rules, and is based on the repetitive routines of initiated persons in the society under description.

(Firth 1950: 183, siteeraus León 2007: 6)

León huomauttaa, ettei Firth vienyt kategorisointia konkreettiselle tasolle (León (2007): 6) ja Firth itsekin piti käsitteen kuvausta keskeneräisenä: ”The technical language necessary for the description of contexts of situation is not developed, nor is there any agreed method for classification.” (Palmer (1968): 177). Määritelmän mukaan konteksti koostuu viestintätapahtuman osapuolista, kielellisestä ja kielenulkoisesta viestinnästä, oleellisista viestintään vaikuttavista esineistä ja tapahtumista sekä kielellisen viestinnän vaikutuksesta vastaanottajaan. (Palmer (1968): 177), Shore (2004): 132). Firth määritteli käsitteen lähtökohtaisesti puhetilanteen kautta: ”In other words, ’meaning’ is a property of the mutually relevant people, things, events in the situation. Some of the events are the noises made by the speakers... and any remoter text which can be apprehended in use can be regarded as having such temporary meaning as is given to it by the reader.” (Palmer (1968): 14). Hallidayn systeemifunktionaalissa ajattelutavassa sekä puheen että tekstin rakenteeseen ja funktioihin kohdistuvia kielellisiä valintoja säätelevää tilannekontekstia määrittävät tekstin ala (field of discourse), tilanteeseen osallistuvien osallistujaroolit (tenor of discourse) ja kielen käyttö tilanteessa, toisin sanoen ’ilmenemismuoto’ (mode of discourse). (Halliday ja Hasan (1985), Shore (2012))

Varsinainen tutkimusaineisto, kokoelma EU-asetuksia, on syntynyt erilaisen prosessin tuloksena erilaisessa tilannekontekstissa kuin sen ensimmäinen verrannaisaineisto, osa Suomen säädöskokoelmaa. EU-aineisto on tuotettu lyhyen ajan sisällä tiukkoja kääntämisen reunaehtoja noudattaen, kun taas SL-aineisto on syntynyt pitkän aikavälin kuluessa Suomen lainsäädännön täydentyessä ja uudistuessa. Ensimmäinen suomenkielinen lakikirja, Ruotsin lain suomennos, ilmestyi jo vuonna 1759 ja Suomen säädöskokoelma sai nykyisen muotonsa teoksina Suomen Laki I ja II vuonna 1955 (Mattila (2010)). Toinen verrannaisaineisto (semfinlex-lait) sisältää Suomen säädöskokoelman (Finlex) alkuperäiset lait vuodesta 1960 aina vuoteen 2018 asti. Kolmas verrannaisaineisto (LaDi) on tekstikokoelma, joka koostuu sekä EU-teksteistä että Suomen säädöskokoelman teksteistä pääosin 2000-luvun alku-

puolelta ja yleiskielinen neljäs verrannaisaineisto (lehdet-osa) sisältää aikakauslehtiä 1990-luvulta ja 2000-luvulta. Aineistot on kuvattu tarkemmin luvussa 4.

2.5 Merkityksen kontekstisidonnaisuudesta

Kollokaatio näyttäytyy abstraktina ilmiönä, joka ei ole tarkasti rajattavissa eikä ase- tu kielenkäyttäjän tietoisien käsityksen piiriin. Jos kääntäjältä kysytään jonkin hänen työkielensä sanan vastinetta toisella työkielellä, saatetaan saada vastaukseksi, että asia riippuu käyttöyhteydestä ja ehkä pyynnön määrittellä tarkemmin, missä yhteydessä sanaa käytetään. Jos taas kysytään kahden sanan, lausekkeen tai sanaliiton vastinetta, autetaan kääntäjää hahmottamaan tilannetta jo vähän paremmin. Vasta kun kääntäjä saa kysytyn termin ympärille enemmän kontekstia, hän voi ehkä tarjota vaihtoehtoisia kohdekielisiä vastineita. Kokemukseni mukaan merkitys välittyy harvoin yksittäisen sanan kautta, ellei sana ole osa suljettua perusleksikkoa, vaikkapa numeraali tai viikonpäivä. Kohdekielisen vastineen tuottamiseksi tarvitaan lisätietoa tekstistä, onko se erikoskielinen tai ammattikielinen, rekisteristä²⁵ tekstin vastaanottajasta, ja niin edelleen. Jos käytettävissä on ympäröivää tekstimassaa, riippuu kääntäjän kokemuksesta ja ammattitaidosta sekä käännoستهتävän luonteesta, miten kollokaation edustumat kääntyvät kohdekielen odotetuiksi vastineiksi. Lukiessamme äidinkielellämme kirjoitettua tekstiä tunnistamme kuitenkin nopeasti, ovatko sen sisältämät kielelliset elementit odotustemme mukaisia.

Firthin teorioita edelleen työstäneen M.A.K. Hallidayn systeemis-funktionaalisisssa kielimallisssa (Halliday (1992), Halliday (1994)) kieli nähdään samoin perusolemukseltaan kontekstin varaan rakentuvana. Merkitys kerrostuu: ”merkityksiä tuotetaan kaikilla kielen tasoilla ja kielen käyttökontekstissa” (Shore (2012): 133). ”If words are not chosen independently by speakers then they convey meaning only as part of the environment in which they are used: they are not meaningful as separate units. Individual words in language production therefore tend to be delexicalized. Delexicalization in Sinclair’s synchronic definition, is the corollary of shared meaning.” (Partington (1993): 186)

Yksittäisillä sanoilla on merkitystä lähinnä siinä suhteessa, miten ne toimivat osana laajempia kielellisiä rakenteita, joissa ne tuottavat osan sen merkityksestä (Palmer (1968): 91). Esimerkiksi käy edellinen virke: suomenkieliselle lukijalle on varmasti selvää, että yksittäisellä sanalla *merkitys* on kaksi erilaista referenttiä, jois-

²⁵Määritelmä: kielimuoto, jonka käyttö kytkeytyy tiettyihin tilanteisiin. Selite: Eri rekisterit voivat poiketa toisistaan mm. sanastollisesti, kieliopillisesti ja syntaktisesti. Rekisterit voivat olla melko suppeisiin tilannekonteksteihin rajattuja (esim. mainonnan kieli, urheilukieli jne.). Tilanteeseen sopivan rekisterin valintaan vaikuttaa mm. tilanteen muodollisuus. Tieteen termipankki 21.05.2018: Kielitiede:rekisteri. (Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Kielitiede:rekisteri>.)

ta ensimmäinen viittaa siihen, miten tärkeänä asiaa pidetään ja jälkimmäinen semantiikan termiin, jolla tarkoitetaan sanan merkityssisältöä.

Erikoiskielissä, ammattikielissä ja niiden käännöstehtävissä lähtökohtana on, että termeillä on mahdollisesti muista vastaavista konteksteista poikkeavia merkityksiä, joista suuri osa on normitettuja tai vähintään konventionaalistuneita (tarkemmin luvussa 3).

2.6 Vapaan ja idiomaattisen valinnan periaatteet

Valitsemme sanamme tapauskohtaisesti kahden erilaisen ja toisiaan täydentävän periaatteen, vapaan valinnan periaatteen (open-choice principle) ja idiomaattisen valinnan periaatteen (idiom principle) mukaisesti (Sinclair (1991), Sinclair et al. (2004)). Idiomaattisen valinnan periaate on kielen oppimisessa hyödyllinen käsite, koska se auttaa hahmottamaan, mikä kielenoppijan tuottamassa tekstissä mahdollisesti on vialla (Herbst (2012): 28).

Vapaan valinnan periaate viittaa sanojen yhdistelemiseen siten, että niiden ensisijaiset merkitykset säilyvät. Yhden kielellisen elementin valitseminen rakenteeseen ei silloin vaikuta toisen tai toisten elementtien valintaan. McKeownin ym. mukaan vapaan valinnan periaate perustuu olettamukseen, että sanat voidaan kuvata kielen yleisten sääntöjen avulla, jotka määrittävät sanojen esiintymistä määrättyssä syntaktisessa rakenteessa yhdessä pääsanansa kanssa. Esimerkiksi *run* + OBJ merkityksessä 'manage' eli 'johtaa' tai 'hallita' olisi vapaan valinnan periaatteen mukainen rakenne. Perusteluna on, että objektin paikalla voi esiintyä melkein mikä tahansa "virtually open-ended" sana, erityisesti kaikki organisaatioita ja yrityksiä yms. kuvaavat sanat (McKeown ja Radeev (1999): 2–3). Samoin substantiivien määreenä esiintyvät yleiset adjektiivit, esimerkiksi *good*, *useful*, *excellent*, kuuluvat vapaiden sanojen yhdistelmiin eivätkä siten teorian mukaan voi olla muodostamassa kollokaatiota. Kollokaation muodostavat osat kuuluvat semanttisesti rajoitetumpaan kategoriaan, esimerkkinä *explode a myth*, jossa verbin *explode* mahdolliset objektit muodostavat rajoitetun joukon sisältäen sellaiset substantiivit kuten *belief*, *idea* tai *theory*, joilla on yhteisiä merkityspiirteitä. McKeownin ym. näkökulmasta katsottuna harvinaiset, mutta systemaattisesti toistensa yhteydessä esiintyvät sanat muodostavat leksikaalisia rakenteita, jotka nousevat tilastollisesti merkitseviksi ja muodostavat siten kollokaatioita. (McKeown ja Radeev (1999))

Kollokaatio toteuttaa idiomaattisen valinnan periaatetta (Sinclair (1991): 115–121). Sitä noudattaessamme yhdistelemme sanoja toisiinsa rajoitetusta, kontekstisidonnaisesta lekseemien valikoimasta. Valikoima voi määräytyä aihealueen perusteella: reaali maailman ilmiöillä ja asioilla, jotka ovat tekemisissä keskenään, on taipumus

esiintyä yhdessä useammin kuin kyseisen tilannekontekstin ulkoisten sanojen kanssa. Firth nimitti tällaisia yhteisesiintymiä **yleisiksi kollokaatioiksi** (common collocations) ja Sinclair näki niiden muodostavan yhteisesiintymisen jatkumon toisen ääripään, jossa frekventit kielelliset yksiköt esiintyvät yhdessä toisten frekventtien yksiköiden kanssa. Vastaavasti R. L. Traskin kielitieteen opiskelijoille suunnatussa sanakirjassa kollokaatio määritellään samaan aihepiiriin kuulumisen perusteella: ”The tendency of certain words to occur together, such as grill (or broil) with meat and toast with bread.” (Trask (1997)). McKeownin ym. määritelmästä poiketen myös kielessä taajaan esiintyvät sanat voivat siis muodostaa yhteisesiintymiä, mutta niiden osien välinen sidos ja muodostumisen motivaatio erottaa ne muista yhteisesiintymien tyypeistä.

Teemme kuitenkin lisäksi paljon sellaisia kielellisiä valintoja, joita fyysinen todellisuus ei luo eikä selitä. Jatkumon toisessa ääripäässä sijaitsevat usein yhdessä esiintyvät yksittäiset leksikaaliset elementit, joiden valikoituminen yhteen johtuu muusta kuin aihealueesta, esimerkiksi konventioista²⁶ tai useammasta kuin yhdestä kielellisestä elementistä koostuvan kielellisen yksikön (linguistic unit) merkityksestä, joka osaltaan muodostuu siksi, että nimenomaan juuri kyseiset kielelliset yksiköt valitaan yhdessä (co-selection). Tällaiset fraasit ymmärretään yhdeksi yksiköksi siitä huolimatta, että ne voivat koostua kahdesta tai useammasta lekseemistä. Idiomaattisen valinnan periaatetta toteuttavat valinnat ovat tavallaan oikopolkuja – kollokaatioiden osat aktivoituvat yhden valinnan kautta ja toimivat kielessä yhtenä kokonaisuutena, joten kielen käyttäjälle jää vähemmän valintoja tehtäväksi. Vaihdamme idiomaattisuuden periaatteesta avoimen valinnan periaatteeseen ja takaisin koko ajan. (Sinclair (1988): 320–323)

Vapaan valinnan periaatteen mukaisesti rakentuva kieli ei yleensä ole kääntäjälle tai maallikolle ongelmallista. Jos syntaktiset rajoitteet sallivat, voidaan sanasanaisella kääntämisellä pärjätä pitkälle. Mitä kauemmas vapaan valinnan periaatteesta liikutaan, sitä vahvemmin idiomaattisen ja äidinkielen lukijan odotuksia vastaavan kielen tuottaminen on riippuvaista kääntäjän kokemuksesta, kielitaidosta ja kielen mekanismien tietoisesta tarkastelusta. Tutkimuksen kohteena oleva substantiivin ja sen potentiaalisten verbikollokaattien muodostamia yhteisesiintymiä on lähtökohtaisesti pidettävä vapaan valinnan periaatteen piiriin kuuluvina, jolloin tutkimukseni tavoitteena on tarkastella kysymystä vapaan ja idiomaattisen valinnan mahdollisesta välitilasta säädöskielessä.

²⁶Konventio: Määritelmä: vakiintunut käytäntö. (Lähde: Tieteen termipankki 2.03.2016: Kielitiede:konventio.)

2.7 Yhteenvedo: kollokaatio

Tämän tutkimuksen kannalta keskeisiä lähtökohtia ovat Firthin ”tilannekonteksti, osakieli, merkityksen hajottaminen eri tasoihin ja merkitysten analyysi systeemeinä ja rakenteina” (osittainen lainaus Shore (2012): 132–134). Tilannekonteksti ja osakieli ovat tutkimukseni aineiston valinnan ja menetelmällisten ratkaisujen taustalla vaikuttavia oletuksia – että kieli on kontekstisidonnaista, että konteksti myös käännöstyössä kolmitasoinen, tekstimassaa laajempi ja abstraktimpi käsite ja että kielen kuvauksen kohteena on autenttinen osakieli. Merkitykseen liittyvät havainnot pyrin näkemään osana kielen leksikaalista rakentumista yhtä sanaa (merkitysyksikköä) suurempina kokonaisuuksina ja merkityksen eriytymisen mekanismina.

Yhteisesiintymän tutkimuksen käytännön valintoja määrittävät seikat liittyvät ajan kulkuun, aineistoon ja tutkimusmenetelmään. Monet tässä luvussa kuvatuista ominaispiirteistä on määritelty etukäteen menetelmän valinnasta johtuen. Etukäteen määritellyt reunaehdot vaikuttavat väistämättä tutkimustuloksiin: esimerkiksi tutkimusikkunan koko määrittää, miten laajalti hakusanan ympäristöä voidaan ottaa huomioon tuloksissa ja tilastollisten menetelmien käyttö vaikuttaa omalta osaltaan tutkimustulosten tulkintaan ja arvioon niiden yleistettävyydestä.

Olen tässä luvussa kuvannut kollokaation ominaisuuksia ja tutkimuskirjallisuudessa esitettyjä näkökulmia ilmiöön. Jotkin ominaisuuksista ovat tämän tutkimuksen menetelmin tarkasteltavissa tutkimusaineistoista tehtyjen havaintojen valossa. Esitän alla listan niistä ominaispiirteistä, joita pidän tässä tutkimuksessa taustaoletuksina yllä kuvatun teoriapohjan perusteella.

Yhteisesiintymä on toistuva -> se on tulevaisuudessa todennäköinen

Yhteisesiintymä voi olla kompositionaalinen

Yhteisesiintymä sallii paradigmaattista vaihtelua

Yhteisesiintymä sallii elementtien välisen etäisyyden

Yhteisesiintymä voi olla konventionaalinen

Yhteisesiintymä voi olla kiinteä

Yhteisesiintymä voi olla sisäkkäinen

Vakiintuma on ei-kompositionaalinen

Semanttinen preferenssi motivoi vakiintuman

Semanttinen prosodia motivoi vakiintuman

Vakiintuma voi sallia paradigmaattista vaihtelua

Vakiintuma voi sallia elementtien välisen etäisyyden

Vakiintuma voi olla konventionaalinen

Kiteymä ei salli paradigmaattista vaihtelua

Kiteymä on normatiivinen

Kiteymä on kiinteä

Kiteymä ei ole sisäkkäinen

Yhdyssana muodostaa poikkeuksen ortografisesta rakenteesta

Seuraavassa luvussa käsittelemme tarkemmin tutkimuksen tekstilajia, suomennettua säädöskieltä tutkimuskohteena ja käännoستهتävänä Euroopan unionin kontekstissa. Tarkastelen myös lyhyesti säädöskielen tutkimusta Suomessa.

Luku 3

Erikoiskieli, säädöskieli, suomennos

Tässä luvussa kuvaan tutkimuksen tekstilajia ja sille tyypillisiä piirteitä. Tutkimuskohteen tekstilaji on säädöskieli ja varsinainen tutkimusaineisto on suomennettua säädöskieltä. Tavoitteeni tässä luvussa on esitellä säädöskieltä tutkimuskohteena, sen tutkimukseen ja kääntämiseen liittyviä kysymyksiä sekä suomennetun säädöskielen erityisominaisuuksia.

Käsittelen ensin erikoiskielen, ammattikielen ja virkakielen käsitteitä ja säädöskieltä niiden edustajana. Tutkimusaineistot ovat lainsäädännön erikoiskieltä, mutta samalla ne edustavat lainkäyttövallan edustajien työssään käyttämiä virkakielisiä tekstejä ja lakien ja muiden säädösten parissa työskentelevien henkilöiden keskenään käyttämää ammattikieltä. Terminologisten ja käännöstieteellisten näkökulmien esittelyn jälkeen tarkastelen EU:n säädöksiä tekstilajinsa edustajana, niiden käännöksiä tutkimuskohteena, sekä yhteisesiintymien rekisterisidonnaisuutta ja lopuksi esittelen säädöskielestä Suomessa tehtyä tutkimusta.

Euroopan unionin tuottamat tekstit voidaan määritellä käännössuomeksi tai toisen tulkinnan mukaan EU:n erikieliseksi samanarvoiseksi versioiksi. Tutkimuksessani pyrin tarkastelemaan tätä kysymystä mikrotasolla leksikaalisten valintojen analyysin avulla.

3.1 Oikeuskieli erikoiskielenä

Tieteen termipankin määritelmän mukaan **erikoiskieli** on ”jonkin tieteen-, ammatin- tai harrastealan kielimuoto, joka eroaa yleiskielestä lähinnä sanastonsa vuoksi”²⁷ Terminologian sanastossa luokittelun perusteena on kielen käyttö tietyllä erikoisalalla.

²⁷Tieteen termipankki 28.05.2018: Terminologiaoppi:erikoiskieli.
(Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Terminologiaoppi:erikoiskieli>.)

Erikoiskielet voidaan jakaa edelleen ammattikieliin, tai vaihtoehtoisen jaottelutavan mukaisesti harrastekieliin tai ammattikieliin, ammattislangeihin, käsityöläiskieleen ja yleistajuistettuun kieleen. Viimeksi mainittu Niemikorven jaottelu heijastaa yleiskielen ja erikoiskielen rajanvedon häilyvyyttä: yleistajuistettu kieli on jo osa yleiskieltä. (Pasanen (2009): 38–39). Erikoiskieli on ”kieli, jota asiantuntijat puhuvat keskenään tai jota asiantuntija puhuu maallikolle.” (Ketola (2002): 22). Yli-Jokipii kuvaa ammattikieltä yksilön näkökulmasta ammattilaisen toiselle alan edustajalle puhumana kielenä: se on ”henkilön työssään käyttämää kieltä” (Yli-Jokipii (2004): 83). Tieteen termipankissa ammattikieli määritellään erikoiskielen alalajiksi.²⁸

Léon pitää käsitettä **osakieli** (restricted language) Firthin yrityksenä kuvata tilannekontekstia kielitieteen keinoin:

With the notion of restricted languages, Firth thought out the different stages of an empirical view of linguistic description, that is the distinction between language of description, language under description, and language of translation, as well as the indeterminacy of language and especially metalanguage. (Léon (2007): 8)

Käytän termistä **restricted language** vastinetta **osakieli** (Shore (2012)), vaikka se siis kattaa vain osan Firthin käsitteestä. Firth käyttää myös termiä **sublanguage** tarkoittamaan jotakin tarkemmin määriteltyä kielenkäytön muotoa, mutta sen suhde termiin ’restricted language’ jää epäselväksi. Erikoiskieli viittaa siten mihi tahansa rajalliseen kieleen, jossa sanasto on laajempi kuin yleiskieli: ”Väitetään, että yleiskieli on osa erikoiskieltä, voidaan tosin ainakin teoriassa perustella sillä, että erikoiskielet ovat sanastoltaan laajempia kuin yleiskieli” (Pasanen (2009): 39). Erikoiskieli on yhteisön joko tietoisesti tai ajan mittaan tiedostamatta yhdessä määrittelemä. Sillä viitataan eri ammattialojen, tieteen, teknologian, politiikan kieliin yhtä hyvin kuin yksittäisen kirjailijan kirjoitustapaan tai henkilön puhetyyliin. Yksittäisellä erikoiskielellä on oma mikrotason kielioppinsa ja sanastonsa, joka on yleiskieltä rajoittuneempi. Firth mainitsee esimerkkinä mahdollisuuden edistää Euroopan yhteisöllisyyttä ja yhteistyötä kiinnittämällä eri osapuolten huomiota tarvittavien kielten hallintaan. (Palmer (1968): 106). Osakieli viittaa sekä tutkimuksen kohteena olevaan kieleen, sen kuvauksessa käytettyyn kieleen että tarpeeseen kääntää kieliä. Se viittaa erilaisten tekstilajien (genre) kieleen, joita ovat muun muassa politiikka, urheilu, kauppa tai teknologia (Firth ja Palmer (1968)). ”Descriptive linguistics is at its best when it concentrates on what I call restricted languages. A restricted language serves a circumscribed field of experience or action and can be said to have its own grammar and dictionary” (Firth 1957, siteeraus Shore (2004)).

²⁸Tieteen termipankki 02.06.2016: Terminologiaoppi:ammattikieli.

Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Terminologiaoppi:ammattikieli.>

Heikki E.S. Mattila on määritellyt **oikeuskielen** yleiskieleen perustuvaksi erikoiskieleksi, jolle tyypillisiä piirteitä ovat kielikohtaiset erityiset termit ja erikoiskieleen oma erityinen kielellinen normi (Mattila (2002): 3). Lait on kuitenkin kirjoitettu yhtä aikaa sekä alan asiantuntijoille että kansalaisille ja ”oikeuskieli poikkeaa yleiskielestä . . . eniten termistönsä eli ammattisanastonsa suhteen” (Mattila (2002): 15). Tekstiä ei sanaston perusteella aina pysty luokittelemaan tietyn alan erikoiskieleksi tai ammattikieleksi, vaan tarvitaan lisäksi tietoa siitä, mihin tarkoitukseen teksti on tehty. Oikeuskieli on ”kielimuoto, jota käytetään oikeudellisessa viestinnässä ja jolla on sanastollisia ja tyylillisiä erityispiirteitä”²⁹. Sen lähisynonyymejä ovat **juridinen kieli**, **lakikieli** ja **säädöskieli**.

Yhteisesiintymä joko merkitystä kantavana yksikkönä tai konventionmukaisena ja odotuksenmukaisena esiintymänä on eräs oikeuskielen ominaisuuksista. Heikki E.S. Mattilan sanoin: ”Kääntämisessäkin inhimillinen kontrolli perustuu viime kädessä kääntäjän yleissivistykseen ja yleistietoon. Tästä syystä hän tarvitsee tietoa oikeuskielen yleisistä ominaisuuksista” (Mattila (2002): 29). Selvää on, että 1990-luvun lopussa koottu tutkimusaineisto ja sen ensimmäinen verrannaisaineisto kuvavat senhetkistä tilannetta ja tutkimuksen painopiste on Euroopan unioniin liittymisen aikaan suomennetun erikoiskielen tarkastelu verrattuna tuolloin voimassa olevaan lainsäädäntöön. Tutkimustuloksia verrataan lisäksi kolmeen muuhun verrannaisaineistoon, joista kaksi edustaa erikoiskieltä ja yksi yleiskieltä. Tarkastelun kohteena on, miten odotuksenmukaisuus toteutuu suomennetuissa teksteissä.

Tarkastelen seuraavassa alaluvussa tarkemmin oikeuskielen lähisynonyymejä, virkakieltä, säädöskieltä ja lakikieltä.

3.2 Virkakieli, säädöskieli ja lakikieli

Kotimaisten kielten keskuksen yhtenä tehtävänä on tutkia virkakieltä ja tarjota neuvontaa siihen liittyvissä kysymyksissä. Myös säädöskieli ja EU-kieli ovat virkakieliä. Kotuksen verkkosivuilta löytyvän määritelmän mukaan **säädöskiehellä** tarkoitetaan ”kansallisen lainsäädännön tekstien, erityisesti lakien ja asetusten kieltä”³⁰ ja viittaankin tällä termillä käsillä olevan tutkimuksen tutkimusaineistoihin: EU-lainsäädäntöön ja Suomen säädöskokoelman sisältämien tekstien kieleen. **Lakikieli** tarkoittaa samaa kuin säädöskieli. Se on vanhastaan tarkoittanut samaa kuin oikeuskieli nykyisin ja sitä käytetään edelleen oikeuskielen synonyyminä.

Säädösteksteissä epäkonventionaalisten tyyli- tai tehostekeinojen käyttäminen ei

²⁹Tieteen termipankki 4.06.2018: Oikeustiede:oikeuskieli.
(Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Oikeustiede:oikeuskieli>.)

³⁰<http://www.kotus.fi/kielitieto/virkakieli/saadoskieli>

ole toivottavaa. Säädöskielessä tulisi yleensäkin välttää monitulkintaisuutta ja pysytellä konventionaalisissa ratkaisuissa, jotka voidaan aina tulkita samoin eri lain-tulkintatilanteissa. Normatiivisuus ja totutuissa rakenteissa ja termeissä pysyminen vahvistavat tulkinnan johdonmukaisuutta, kun on voitava olettaa, että termeillä viitataan kaikissa lainkohdissa samoihin tarkoitteisiin. Toisaalta säädöskielessä tava-taan osittaista polysemiaa, jossa yhdellä termillä voidaan viitata useaan tarkoittee-seen (Mattila 1997, viittaus teoksessa Vanhatalo (2003): 24).

Työlainsäädännön tekstejä systeemis-funktionaalisen kielimallin rekisteri- ja teks-tilajiteorian viitekehyksessä tutkineen Ketolan tulosten perusteella voidaan todeta, että työlainsäädännön tekstit ovat abstrakteja (verbin merkitys selviää sen komple-mentista; tekijöinä on abstrakteja kollektiivisia yhteisöjä), ne ovat staattisia (staattisia verbejä on runsaasti; toiminta kohdistuu usein asiakirjoihin; prosessit ovat useim-miten eksistentiaalisia), ne ovat ohjailevia (säännöt, sanavalinnat) ja eksplisiittisiä (asioita määrittellen tarkasti) (Ketola (2002): 154). Lakikieli muuttuu kuitenkin yhteiskunnan muuttumisen myötä ja yhä moninaisempia tilanteita koskevia lakeja säädetään. Lakikielessä kuvataan entistä abstraktimpia asioita, joita voidaan lain tulkintatilanteissa helpommin yleistää. Kielen muuttumisen lisäksi sen välittämän informaation määrä kasvaa (Ketola (2002): 20–21). Pasasen mukaan lain ja hallin-non kieli on kulttuurisidonnaista erikoiskieltä, johon ei välttämättä päde yleissääntö kansainvälisen erikoisalan taipumuksesta siirtää käsitesisältö muuttumattomana yli kieli- ja kulttuurirajojen. Siten sanojen merkitys riippuu aina kontekstista. (Pasanen (2009): 40)

Euroopan unionissa suomennettujen asetusten aineisto edustaa säädöskieltä, joka on oikeustieteen alaan kuuluvaa erikoiskieltä. ”The restricted languages of official translation, in spite of careful control and attempts at valid equivalents deserve detailed study” (Firth 1957: 87, siteeraus Léon (2007): 12). Tämä tutkimus tuo leksikaalisen näkökulman Firthin osakielten tutkimukseen.

3.3 Terminologia näkökulmia

Erikoiskielen leksikon kartoituksessa terminologiset näkemykset ovat väistämättä keskeisiä. Kieltä käyttävä yhteisö määrittelee terminsä käyttääkseen niitä syste-maattisesti viestinnässään. Perinteisessä termityössä tavoitteena on suhde termin ja käsitteen välillä (Pearson (1998): 15) ja tästä suhteesta tulee pysyvä käyttäjien yksimielisyyden kautta. Yksimielisyys syntyy asiantuntijoita konsultoimalla, jolloin tuloksena on erikoiskielessä hyväksytty termi. Termityötä tehdään myös pragmaatti-semmista lähtökohdista tutkimalla erikoisalan kieliaineistoja lähtien liikkeelle kielen käyttäjien kokemuksista (Pearson (1998): 4, 16). ”Words become terms, i.e. acquire or have protected status when they are used in special subject domains” (Pearson

(1998): 7). Jotkin termeistä ovat hyvinkin harvinaisia, eikä niitä käytetä yleiskielessä, kun taas toisilla voi olla yleiskielisen merkityksen lisäksi erikoisalakohmainen erityismerkitys terminä (Pearson (1998): 27)

Perinteisissä terminologian alan teorioissa termi on tavallisesti kontrastoitu yleiskielen sanan kanssa siten, että termi on alalla yhteisesti sovittu nimike käsitteelle, jolla on yksiselitteinen tarkoite ja jonka dokumentaatiossa käsite, tarkoite ja termi tulevat kaikki määritellyiksi osana termin kuvausta. Pearson toteaa kuitenkin, että termille ei ole yleispätevää määritelmää. Tämä johtuu osaltaan siitä, että termit ovat muodoltaan saman näköisiä kuin (tavalliset) sanat, eikä mikään yhteisö pysty tuottamaan yhteisesti sovittuja määritelmiä kaikille niille sanoille, jotka toimivat kyseisen alan teksteissä tai puheessa termien tavoin. Pragmaattisten teorioiden mukaan yhteisö eli termin käyttäjät määrittelevät itse, mitä termi tarkoittaa. Pearsonin ratkaisu on etsiä termejä teksteistä kontekstin perusteella, joka hänen tutkimustulostensa mukaan paljastaa termin. (Pearson (1998): 7–8). Vaikka termit olisi hyvin dokumentoitu termipankkeihin tai sanalistoisiin, niiden käyttäjät eivät välttämättä aina käytä niitä systemaattisesti, vaan saattavat käyttää niitä uudessa merkityksessä (Pearson (1998): 15) tai ne voivat olla polyseemisiä (Pearson (1998): 20). Samat termit voivat siten olla eri alojen erikoiskielissä käytössä toisistaan poikkeavissa merkityksissä ja yleiskielen sanat voivat saada erilaisia tarkemmin rajattuja merkityksiä eri erikoiskielissä. Tavallinen esimerkki löytyy tietotekniikan piiristä, jossa **hiirtä** napautetaan erilaisten **ikkunoiden** avaamiseksi. Lisäksi sanakirjoihin, termipankkeihin ja käännosmuisteihin voidaan tallentaa vain osa oleellisistakin termeistä – vain ydin (henkilökohtainen tiedonanto Stina Ojala 5.2.2017).

Käytännössä erikoiskielen käyttäjät sopivat vakiintuneista termeistä, jotka kattavat alan keskeisiä käsitteitä. Kieltä käyttävässä yhteisössä niistä vallitsee yhteisymmärrys. Termityön tuloksena alan keskeisiä käsitteitä kattavat termit kirjataan termipankkeihin, sanakirjoihin tai sanastoihin, mutta koska prosessi vie aikaa, ovat termit niihin päätyessään jo vakiintuneita (Pasanen (2009): 37). Perinteisen terminologian tavoittelemia ihannetermejä ei todellisessa kielenkäytössä juuri ole. Ihanteellinen termi on yksimerkityksinen, se ei riipu kontekstista eikä sillä ole synonyymejä (Pasanen (2009): 29–30).

Termityössä termit viittaavat useimmiten yksittäisiin käsitteisiin, joilla on yhteisesti sovittu nimike. Kaksiosaisen termin, esimerkiksi teknisen yhteisesiintymän (technical collocation) tarkoitteet ovat tilanteita tai propositioita, jotka koostuvat enimmäkseen useista toisiinsa liittyvistä käsitteistä. (Wright (1997): 13–24). Molempien tarkoitus on sama: ”to firmly delineate the conceptual and linguistic space occupied by strictly defined individual or closely associated concepts and units of knowledge” (Wright (1997): 16). Yhteisesiintymä tarkoittaa kiinteää lauseketta (*set phrase*, *phraseological unit*), jonka osat esiintyvät usein yhdessä, mutta eivät vält-

tämättä peräkkäin, esimerkiksi *file a patent*. ”This principle of potential dislocation within the stream of discourse constitutes the essential property of combinatory forms, hence of ”co-location” (Wright (1997): 15). Toisinaan monisanaisia ilmauksia pidetään moniosaisina termeinä tai kiinteinä fraaseina puhtaasti sen vuoksi, että ne esiintyvät niin usein yhdessä, mutta aina ei ole selvää, onko tällainen yhdistelmä erillinen termi. Wrightin mukaan kääntämistä voidaan pitää testinä: jos termi kääntyy toiseen kieleen yhdeksi sanaksi tai yhdyssanaksi, voidaan olettaa että kyseessä on lähtökieleessä kiinteä lauseke. (Wright (1997): 15)

Yleiskielessä taaajaan esiintyvät sanat eivät Hallidayn mukaan voi olla kollokaatioiden osia, paitsi erikois- ja ammattikielissä. Siten esimerkiksi englannin *file* yhteisesiintyy tietoteknisissä teksteissä verbien *delete* ja *create* kanssa, jolloin sen merkitys poikkeaa yleiskielisestä merkityksestä. (McKeown ja Radeev (1999)). Sinclair otti myös kantaa kysymykseen yhteisesiintymän osien esiintymistajuuksista suhteessa toisiinsa (ks. vastavuoroisuus alaluvussa 2.2.3) ja eräs tämän tutkimuksen tavoitteista on tarkastella tätä kysymystä aineiston valossa.

Käytän tutkittavasta substantiivista nimitystä **termi**. Pidän niiden luokittelua termeiksi perusteltuna, koska se on dokumentoitu EU:n termipankkeihin (Eurodicautom, IATE). Nojaudun tässä perinteisen terminologian kriteereihin termin statuksen syntymisestä. Tavoitteenani on myös tarkastella termien systemaattista käytön toteutumista tutkimusaineistoissa (Pearson (1998)).

3.4 Käännöstieteellisiä näkökulmia

Korpuspohjainen käännöstieteellinen tutkimus ja kiinnostus käännöskielen tutkimukseen on 1990-luvun alun jälkeen lisääntynyt nopeasti (Bernardini (2015)), kun ”Translation scholars grew interested in the set of ’fingerprints’ that one language leaves on another when texts are translated between the two ...” (Gellerstam (1996), siteeraus Bernardini (2015): 515). Universaaliteorioista (Baker (1996)) huomio on siirtynyt tendenssien tutkimuksen puolelle ja samalla kieliaineistojen laatuun, edustavuuteen ja vertailukelpoisuuteen liittyvät kysymykset ovat nostaneet päätään (Bernardini (2015): 517).

Käännössuomen valinta tutkimusaineistoksi liittyy tutkimuksen polysysteemi-teorian viitekehykseen, jossa käännöskieli nostetaan tutkimuskohteiden joukkoon ja joka näin motivoi aineiston kokoamista käännettyistä teksteistä ja niiden vertaamista alun perin kohdekielellä kirjoitettuihin teksteihin (Kenny (2001): 49–50, Hermans (1999)). Tässä tutkimuksessa käännetty säädösteksti oletetaan käännöskielen systeemiiksi, jolla on paikkansa suomalaisen säädöskielen rinnalla. Käytännössä nämä kaksi systeemiä sekoittuvat säädöskielen kokoelmaksi, jossa niillä on yhdenvertainen ase-

ma Suomen oikeudellisessa kentässä ja yhtä suuri painoarvo säädösten tulkinnessa. Käännössuomesta kielimuotona ja polysysteemiteoriasta löytyy selkeä katsaus Jarmo H. Jantusen synonymiaa käännössuomessa käsittelevästä tutkimuksesta (Jantunen (2004): 33–39). Käännöstieteen piirissä on myös laajasti tutkittu normeja³¹ ja konventioita³² ja niiden roolia käännösstrategioiden valinnassa (Hermans (1999), Toury (1995), Toury (2012), Chesterman (1997), Schäffner (1997), Genzler (2001)).

Ekvivalenssin toteutumista kirjallisuuden käännöksissä kriittisesti tutkinut Toury (Toury (1995), Toury (1999), Toury (2012)) kohdisti huomionsa lähdetekstin ensisijaisuuden sijaan lähdetekstin ja käännöksen suhteeseen ja yhteiskunnan tai kääntäjän valintoja heijasteleviin normeihin. Genzlerin mukaan ”The eventual goal of Toury’s theory was to establish a hierarchy of interrelated factors (constraints) that determine (govern) the translation product. In short, Toury demanded that translation theory include cultural-historical ‘facts’, a set of laws that he calls ‘translation norms’” (Genzler (2001): 127). Touryn normit voivat toimia eri aikakausina eri tavoin, ristiriitaisia ja muuttuvia ja ne vaikuttavat käännöstoiminnan kaikilla tasoilla: **alkunormit** (initial norms) kattavat strategiset päätökset, joita kääntäjä tekee kotoittamisen ja vieraannuttamisen suhteen, toisin sanoen hän ratkaisee hyväksyttävyyteen ja adekvaattisuuteen liittyviä kysymyksiä. **Ennakkonormit** (preliminary norms) ovat kääntäjän toiminnan ulkopuolisia tekijöitä ja heijastavat kyseisen yhteisön käännösstrategiaan vaikuttavia tekijöitä, esimerkiksi mitä tekstejä käännetään ja mitä jätetään kääntämättä. **Toimintanormit** (operational norms) vaikuttavat yksityiskohtia koskevien päätösten tasolla itse käännöstapahtumassa ja niiden joukkoon lukeutuvat leksikaalisia valintoja ohjaavat lingvistiset normit (textual-linguistic norms), lähdetekstin kielellisen materiaalin korvaaminen kohdekielisellä materiaalilla. Ei voida olettaa, että kohdekielen ja käännöskielen tuottamista ohjaisivat samat normit. (Toury (2000): 202 ja Toury (2012): 81–85)

Käännösstrategia on kääntäjän tietoista valintaa niiden normien ja konventioiden välillä, joita lähtökulttuurin ja kohdekulttuurin/tulokulttuurin järjestelmissä on vallalla (Chesterman (1997)). Kääntäjä joutuu valitsemaan, minkä järjestelmän normien mukaan hän toimii: jos valinta on mukautuminen kohdekulttuurin/tulokulttuurin normeihin, hän ”täyttää tulokulttuurin kielellisten normien ja tekstikonventioiden vaatimukset”³³ ja täyttää siten hyväksyttävyyden vaatimuksen. Käännös on tuolloin adekvaatti, eli se noudattaa tulokielen normeja ja siinä on otettu huomioon käänn-

³¹Tieteen termipankki 02.06.2016: Kielitiede:normi. Määritelmä. 1. puheyhteisön jäsenten kielitajun piirre eli luonnollinen normi, joka ohjaa kielellistä käyttäytymistä. 2. kielen standardimuodon eli kirja- tai yleiskielen sääntö, keinokeinen normi.

(Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Kielitiede:normi>.)

³²Tieteen termipankki 02.06.2016: Määritelmä. vakiintunut käytäntö. Kielitiede:konventio. (Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Kielitiede:konventio>.)

³³Tieteen termipankki 02.06.2016: Kielitiede:hyväksyttävyyys.

(Tarkka osoite: <http://www.tieteentermipankki.fi/wiki/Kielitiede:hyväksyttävyyys>.)

nöksen lukijoiden tarpeet ja odotukset³⁴ (Toury (2012): 79–81). Konventiot, normit ja säännöt määrittelevät kääntäjän toimintaa ja ohjaavat käännösstrategian valintaa. Säännöt ovat näistä velvoittavimpia ja konventiot heikommin kääntäjää sitovia ilmiöitä. Rajanveto on vaikeaa, koska normien lisäksi myös säännöt muuttuvat ja niitä muutetaan tarpeen mukaan ajan myötä. Kääntäjä ei aina havaitse yhteisesiintymää lähtökielessä ja kääntää sen elementit kompositionaalisina erikseen. Lähtökielen konventiot saattavat tällöin siirtyä käännökseen ja aiheuttaa odotustenvastaisen ilmauksen, joka on mahdollinen kielessä, mutta ei tyypillinen tai tavallinen.

Kääntäjä valitsee käännösstrategian myös kääntäessään kollokaatioita, tietoisesti tai tiedostamatta. Esimerkiksi yhdyssanan määriteosaksi voi valikoitua suora sitaattilaina tai suora käännöslaina. Tietoiset strategiset valinnat kohdekielen ja sen kulttuurin tai lähtökielen ja sen kulttuurin painottamisesta käännöstyössä sijoittuvat Touryn normihierarkiassa toimintanormeihin ja nämä valinnat määrittävät leksikaalisten valintojen normien aktivoitumisen tekstin tasolla.

Tämän tutkimuksen aineiston tuottaneiden kääntäjien valitsemien mahdollisten käännösstrategioiden sijoittumisesta Touryn normihierarkiaan voidaan tehdä seuraavia tilannekontekstiin liittyviä huomioita: ensinnäkin nimenomaan asetusten kääntämiselle ei Suomen liittyessä Euroopan yhteisöihin ollut vaihtoehtoja (käännöstehtävän toteuttamistavalle olisi saattanut ollakin), yhteisöjen lainsäädäntö eri kielillä katsottiin erikielisiksi versioiksi mieluummin kuin käännöksiksi, joten tarkkaan ottaen käännöksiä ei edes ollut olemassa. Käännöstehtävää tukevat normit ovat siten määräytyneet muualla kuin kääntäjien työpöydällä. Yksityiskohtien tasolla EU-asetusten kääntäjä on sidottu käännöstoiminnan sääntöihin, toimintaohjeisiin ja termipankkeihin. Euroopan yhteisöjen asetusten kääntäjä ei siis oletettavasti ole soveltanut Touryn tarkoittamassa mielessä kohdekulttuuria painottavia käännösstrategioita (Toury (2012): 79). Kun kääntäjät systemaattisesti valitsevat samoja käännösvastineita, on mahdollista, että tuloksena on ajan mittaan uusi kohdekielen muoto (*a distinct variety of TL*) (Toury (2012): 243).

Yhteisesiintymien tunnistaminen ja tietoisuuden lisääminen auttavat kääntäjää tuottamaan luonnollista, idiomaattista ja norminmukaista kieltä. Yhteisesiintymät usein erityisiä vaikeuksia käännettäessä äidinkielestä vieraan kielen suuntaan, koska vieraille kielelle tyypillisten yhteisesiintymien arvaaminen on tavallisesti mahdotonta (Heid (2001): 788). Yhteisesiintymät ovat äidinkieltään käyttävälle läpinäkyviä niiden tulkinnan puolella, mutta tietoinen tuottaminen ilman tilannekontekstia tai kontekstia tuottaa vaikeuksia (Heid (1999)). Tulkinta ja luokittelu hyväksyttäviin yhdistelmiin ei aina äidinkieliseltäkään kielen puhujalta onnistu sujuvasti. McKeown ym. raportoivat Bensonin 1980-luvulla tekemistä kyselyistä, joissa yhteisesiintymien

³⁴Tieteen termipankki 02.06.2016: Käännöstiede:adekvaattisuus.
(Tarkka osoite: [http://www.tieteentermipankki.fi/wiki/Käännöstiede:adekvaattisuus.](http://www.tieteentermipankki.fi/wiki/Käännöstiede:adekvaattisuus))

idiomaattisuutta kysyttäessä 24 prosenttia brittiläisten ja amerikanenglannin äidinkielisten puhujien arvioista osui oikeaan (McKeown ja Radeev (1999): 4). Puhuttaessa tai kirjoitettaessa yhteisesiintymät useimmiten tuottuvat idiomaattisiksi. Voidaankin kysyä, voisiko ammattikielten ja erikoiskielten kääntäminen toimintana silttenkin muistuttaa enemmän vieraan kielen tuottamista kuin äidinkielen tekstin tuottamista.

Hallinnollisen kielen käännösten tavoitteena on selkeä ja idiomaattinen kieli, mutta myös mahdollisimman tarkka lähtökielisen ilmauksen merkityksen siirtäminen kohdekieleen. Toisinaan molemmat eivät voi toteutua yhtä aikaa. Lähdekulttuurin ilmiöille ei aina ole vastineita kohdekulttuurissa, jolloin yhteisesiintymille ei välttämättä ole kohdekielellä odotuksenmukaista vaihtoehtoa. Useimmat kääntäjän käännösstrategian valinnoista eivät ole tietoisia. (Baker (2011): 58–67). Kuten Mattila huomauttaa, Euroopan unionin termien kääntämisessä on kääntäjän valittavana useita tietoisiaakin käännösstrategioita. Näitä ovat 1) suora sitaattilaina, 2) osittainen sitaattilaina, 3) suora käänöslaina, 4) rakennelaina ja 5) yksinkertaistava käänös (Mattila (2002): 195–196). Joistakin käänösvastineista on Suomessakin käyty neuvotteluja, yleensä EU:n ydinsanastosta, kuten Euroopan yhteisöjen nimitys Euroopan **unioni**.

Käännösten voidaan myös haluta näyttävän käänöksiltä, erilaisista syistä. Suomalaisessa oikeusjärjestelmässä toimivien voi olla hyödyllistä nähdä tekstistä, että se on osa Euroopan unionin lainsäädäntöä kansallisten säädösten sijaan. EU-kieli on tunnustettu omaksi kielivariantiksi (Koskinen (2000a)) ja Kotimaisten kielten keskuksessa toimii erityinen EU-kielenhuoltaja. Tässä mielessä Venutin (Venuti 1995, viittaus teoksessa Koskinen (2000a): 52–53) vieraannuttava käänösstrategia väljästi tulkittuna tarkoittaisi selkeästi ilmipantua viestiä EU-kulttuurin ja yhteisöjen tavoitteiden ja tarpeiden näkökulmasta. EU-lainsäädännön piiriin selvästi kuuluva säädös ohjaa lain tulkitsijaa lukemaan tekstin EU-kielenä ja suodattamaan sen käsitteet ja termit EU-kielen tulkintamekanismin läpi. Vieraannuttamisen käsite on kiinnostava EU:ssa toimivan kääntäjän kannalta siinä mielessä, että hänelle EU-kulttuuri ja EU-säädösten suomi saattavat ajan mittaan tulla tutummiksi ja kansallinen säädöskieli voi alkaa edustaa vähemmän tuttua kieltä.

Sinclairin leksikografian puolella tekemä työ (erityisesti korpuslingvistiin menetelmiin perustuva sanakirja COBUILD) toi korpuslingvistiikan menetelmät ja ajatusmallit paitsi käännöstieteen tutkijoiden, myös kääntäjien tietoisuuteen. Normatiivisten ja preskriptiivisten suhtautumistapojen haastaminen ja autenttisen kielen nostaminen tutkimuksen kohteeksi ja kielen luonnollisuuden ja odotuksenmukaisuuden arvioinnin sitominen todelliseen kielenkäyttöön on tuonut kääntäjän työhön uusia työkaluja. Autenttisen kielen tarkastelu työhön luonnollisena sisältyvänä tehtävänä on nykyisin arkipäivää, kun erilaiset sovellukset tarjoavat tietoa helposti ja

nopeasti yksittäisenkin käännöspulman ratkaisun tueksi.

3.5 Autenttinen erikoiskielen aineisto

Aineiston autenttisuus tarkoittaa kielentutkimuksen piirissä sitä, että tutkimusaineisto sisältää todellista kieltä, puhetta, viittomia tai tekstiä. Tarkastellaan sitä, mitä kielessä tapahtuu sen sijaan että keskityttäisiin siihen, mikä on mahdollista. Jos seuraamme J. R. Firthin ajatusta ‘renewal of connection’ (Palmer (1968)), pyrimme palauttamaan yhteyden kielen tutkijan ja todellisen kielen välillä ja tavoittelemme tutkittavan lausuman merkityksen kuvausta mahdollisimman kattavasti. Tässä tutkimuksessa autenttista aineistoa päästään tarkastelemaan vain hyvin pienessä mittakaavassa manuaalisen työn suuren osuuden vuoksi. Silti tavoitteenani on toteuttaa Firthin näkemyksiä mahdollisimman pitkälle.

Valitaan tutkimuskohteeksi jokin rajattu kielen osa-alue, sen osakieli (sublanguage), jolloin noudatetaan Firthin neuvoa: ”it is unnecessary, perhaps inadvisable to attempt a structural and systemic account of a language as a whole” (Palmer (1968): 200). Yhteisesiintymät kartoitetaan ensin systemaattisesti erikoiskielen tutkimusaineistosta valittujen ilmausten osalta, minkä jälkeen voidaan laatia täydellinen lista yhteisesiintyjistä. Jokaista tutkittua sanaa kohden oletetaan löytyvän kollokaatioita (groups of collocations), joiden avulla tunnistetuille merkityksille voidaan ryhtyä antamaan nimikkeitä ja määritelmiä (Palmer (1968): 181). Käsillä olevan tutkimuksen aineisto on suomennettua erikoiskieltä, ja tutkimuksen tavoitteena on pyrkiä ensin määrittämään tutkittava elementti mahdollisimman aineistolähtöisesti ja sen jälkeen löytämään yhteisesiintymiä, joita verrataan alun perin suomeksi kirjoitettuja tekstejä sisältävään verrannaisaineistoon.

Euroopan unionin lainsäädäntö erikoiskielen edustajana

Euroopan unionin toimielinten tuottamien tekstien suomennokset ovat EU-kieltä tai EU-suomea. Kysymykseen siitä, onko EU-kieli tai yleisemminkin kansainvälisten organisaatioiden tai instituutioiden käyttämä kieli erikoiskieltä löytyy erilaisia näkökulmia. Trosborg pitää EU-kieltä hybridikielenä, jossa on ominaisuuksia sekä kansallisista kielistä että omasta EU-kantaisesta aineksestä (Trosborg (1997)). Alarotu pohtii yleiskielisyyden ja EU-kielen rajapintaa ja roolia EU-termien esiintymistä lehtiteksteissä koskevassa tutkimuksessaan ja toteaa, ettei EU-kielen määrittely erikoiskieleksi ole yksiselitteistä tai kiistatonta (Alarotu (2001): 31). EU:n tuottama valtava tekstimassa sisältää kuitenkin minkä tahansa yhteisön tavoin monenlaisia kielivariantteja edustavia ja monia eri tarkoituksia varten julkaistuja tekstejä

– markkinointimateriaalista lainsäädäntöön. Yhdessä tekstissä voi olla monen eri alan erikoiskielen vaikutusta, esimerkiksi EU-rahoitusmahdollisuuksien markkinointitekstistä saattaa löytyä kaupan, tekniikan, maatalouden tai yhteiskuntatieteiden erikoiskielten terminologiaa ja rakenteita. Kun lisäksi katsotaan, että säädöskielen tulisi olla ymmärrettävää ja selkeää, on sen sisältämän yleiskielen osuuden määrä ratkaisevassa asemassa, kun tavoitellaan kansalaisten kokemusta kielen tuttuudesta ja ymmärrettävyydestä. Kun vielä otetaan huomioon Trosborgin ajatukset monikielisen ja -kulttuurisen yhteisön hybridikielestä, on EU-tekstien sijoittaminen kansallisiin tekstilajeihin ja tekstityyppeihin kuuluvaksi haasteellista.

Euroopan yhteisöjen ja myöhemmin Euroopan unionin liittymisvaiheessa ja hyvän aikaa sen jälkeen 1990-luvun lopulla ja 2000-luvun alkuvuosina suomalaiset kohtasivat käännösten muodossa tekstimassan, jonka erilaisuuteen tuttuihin tekstilajeihin, viranomaisteksteihin ja virkakieleen verrattuna ei jälkikäteen arvioituna olisi voitu mitenkään varustautua. Kansalaisten reaktiot ”EU-jargonian” kohtaan olivat paikoitellen hyvinkin kärkeviä aikana, jolloin sosiaalinen media ei vielä ollut avoin kaikille mielipiteille tämän päivän tapaan. Arvostelevat lausunnot saivat huomiota, kun ne julkaistiin päivälehtien mielipidekirjoituksina. Pirjo Karvosen Euroopan yhteisöjen tekstien ymmärrettävyyttä kartoittava tutkimus toi vuonna 1997 keskusteluun kielenhuoltajan näkökulman (Karvonen (1997)).

Tarkastelen tässä tutkimuksessa kysymystä, tukeeko yhteisesiintymien analyysi näkemystä, jonka mukaan Euroopan unionin tuottama lainsäädäntö muodostaa oman hybridikielen, EU-asetusten kielen (Trosborg (1997), Karvonen (1997), Probirskaja (2009)).

Käännöstoiminta EU:ssa ja erikieliset versiot

Varsinainen tutkimusaineisto on käännöskieltä. Se on tuotettu Suomen Euroopan yhteisöihin liittymisen aikaan 1990-luvun alkupuoliskolla. Tavoitteena tuolloin on ollut kääntää siihen asti säädetyt asetukset ja direktiivit mahdollisimman nopeasti ja päästä uuden säädösmateriaalin tuottamisen kanssa samaan sykliin vanhojen jäsenmaiden kanssa.

Vertaan tutkimusaineistosta tekemiäni havaintoja verrannaisaineistoon, joka sisältöltään vastaa mahdollisimman hyvin asetusten aihealueita. Tekstit ovat Suomen säädöskokoelmasta, jonka säädökset on 1800-luvun loppuun saakka suomennettu ruotsin kielestä. Käännettyjä lakeja ei kuitenkaan sisälly toisena verrannaisaineistona toimivaan säädöskokoelman osakorpuksen.

Euroopan yhteisön (nyk. Euroopan unioni) lainsäädännön käännöshankkeelle ominaista oli uusia käsitteitä tarkoittavien termien puute ja kiire saada käännös-

työ nopeasti valmiiksi, erityisesti mittavan käännöstyön alkuvuosina Suomen liityessä Euroopan yhteisöjen jäseneksi vuonna 1995. Muutaman vuoden aikana EU:n termipankkiin Eurodicautomiin³⁵ koottiin noin 100 000 suomenkielistä termiä täydentämällä termipankkia kääntämällä siinä jo aikaisemmin ollutta aineistoa (Alarotu (2001): 38). Kääntäjät olivat sidottuja käyttämään Eurodicautomiin tallennettuja termejä, joiden vaihtaminen mahdollisesti parempiin tai jo aikaisemmin käytössä olleisiin oli melko mahdotonta. Termien päivittäminen Eurodicautomiin oli jatkuva prosessi. Vuonna 2004 yhteisössä siirryttiin uuteen IATE³⁶-termitietokantaan. Myös Euroopan unionin käännoyksiköiden sitoutuminen lähtökielen konventioihin erikielisten versioiden päivittämisen ja toisiinsa vertaamisen helpottamiseksi on alkuaikoina aiheuttanut kääntäjälle päänvaivaa (Karvonen (1997)). 1990-luvun loppupuolella Systran-konekäännintä hyödynnettiin vain Euroopan unionin isojen kielten välisissä käännoyksissä (Mattila (2002): 539), eikä sen ottaminen käyttöön uusien jäsenmaiden pääkielten rakenteista poikkeavien kielten kääntämisessä osoittautunut kannattavaksi (Mattila (2002): 530). Myös Kaisa Koskinen erottaa Euroopan unionin säädösten kääntämisen omanlaiseksi käännoystehtäväkseen EU:n kulttuurin sisäisen ja ulkoisen kääntämisen lisäksi (Koskinen (2000a): 87).

Tutkimusaineiston asetusten lähtökielestä tai lähtökielistä ei ole varmuutta, vaikka valistuneita arvauksia voi tehdä – todennäköisimmin lähtökieli on ollut yksi EU:n isoista kielistä (englanti, ranska tai saksa). Tekstin on voinut kääntää yksi tai useampi anonyymi kääntäjä ja hänen työparinaan oikolukijana toiminut toinen yhtä anonyymi kääntäjä. Tekstejä tai tekstifragmentteja käännettiin, muokattiin ja käännettiin uudelleen kaikille virallisille jäsenmaiden kielille tavoitteena erikieliset ”versiot”, eivät ”käännoykset”. Alkuperäinen teksti hävisi käytännössä eikä sitä enää versioiden valmistuttua ollut olemassa, koska millekään kielelle ei ollut tarkoituksenmukaista antaa ensisijaista asemaa. Tavoitteena ja oletuksena oli eri versioiden välinen sisällöllinen ekvivalenssi: kielellinen ja muodollinen vastaavuus. (Koskinen (2000a): 84–85) Tiedossani ei ole, millaisia apuja yhteisöjen termipankissa IATEssa (aineiston kokoamisen aikaan Eurodicautom) kääntäjälle löytyi aineiston käännoistyön aikaan 1990-luvun keskivaiheilla. Komission käännoyksikössä oli tuolloin kääntäjien tukena erilaisia sanalistoja, suosituksia ja velvoittavia ohjeita, mutta niiden laajuus ei voinut olla tekstien määrä huomioiden kovinkaan merkittävä. Kollektiivisesti organisoituneessa käännoyskulttuurissa yksittäisen kääntäjän rooli uusien termien luomisessa on ollut rajallinen, ellei olematon (Koskinen (2000b)).

Kääntämisen käytänteitä sovellettiin liittymisvaiheen ja massiivisen käännoysurakan aikana kaikkiin kieliin käännettäessä ottamatta huomioon kielikohtaisia erityispiirteitä. Kotimaisten kielten keskuksen EU-kielenhuoltaja Aino Piehl kirjoittaa seu-

³⁵www.eur.int/, elokuu 2004

³⁶<http://iate.europa.eu>

rauksista näin:

”Säädösten kääntämistä on kahlehtinut ns. pistesääntö, joka on vaatinut, että virkkeitä tulee jokaisella kielellä olla yhtä monta. Niinpä virkkeeseen on ison kirjaimen ja pisteen väliin täytynyt joka kielellä sisällyttää sama asiamäärä. Näin on voitu kätevästi viitata säädöksen tiettyyn virkkeeseen kielestä riippumatta. Asia ei kuitenkaan jakaudu luontevasti kaikilla kielillä virkkeisiin samalla tavalla. Vaikka suomen kieli on kuuluisa lauseenvastikkeistaan, se ei tarjoa ranskan tai englannin veroista valikoimaa keinoja, joilla asia toisensa perään saadaan liitettyksi lauseeseen ilman, että niiden sanomiseksi tarvitaan sivulause tai uusi virke. . . . Olisi toivottavaa, että kääntäjät ja lingvistijuristit voisivat vapaammin sopeuttaa tekstinsä oman maansa tekstikulttuuriin. Kenties tämä huomataan välttämättömäksi, kun unioniin kohta tulee kymmenen uutta virallista kieltä. Myös jäsenmaissa on oltava valmiita antamaan tukea ja auttamaan uudistuspyrkimyksissä. Jotta päästäisiin lähemmäs kansalaisia, olisi EU:n tekstinikin tunnettava kaikissa jäsenmaissa omilta.” (Piehl (2003))

Vaikka EU:ssa on kuluneina vuosikymmeninä alettu kiinnittää huomiota muun muassa selkokielisyyteen, ovat 2000-luvun ensimmäisen vuosikymmenen näkemykset EU:n käännöstötoiminnasta, kääntäjän roolista ja käytettävissä olevista käännösstrategioista tutkimukseni kannalta edelleen relevantteja.

Euroopan unionin käännöksiin kohdistui 1900-luvun lopulla kovaakin kritiikkiä (Karvonen (1997)), mikä osaltaan johtui kääntämisurakan laajuudesta ja työtä määrävistä tiukoista ehdoista, muun muassa virkerakenteen yhdenmukaisuuden vaatimuksesta eli yllä mainitusta pistesäännöstä. Kysymys käännöstötoiminnan käytännön järjestelyistä tai niiden onnistumisesta ei kuitenkaan ole tutkimukseni kannalta keskeinen: kysymys on siitä, mitä valintoja kääntäjä tekee vallitsevissa olosuhteissa ja miten nämä valinnat mahdollisesti vaikuttavat tulevaisuudessa säädöskielen termistöön. Tutkimuksen avulla voidaan arvioida, onko yhteiseisintyminen ilmiönä tärkeää tuoda kääntäjien ja lainsäätäjien tietoisuuteen ja koulutukseen.

EU-lainsäädännön integrointi Suomen säädöskieleen

Asetukset ovat silloisen Euroopan yhteisön ja nykyisen Euroopan unionin yhteistä oikeutta. Ne säädetään EU:n yhteispäätösmenettelyllä (nykyisin ns. tavallinen lainsäätämismenettely). Euroopan komission tehtävänä oli laatia säädösehdotus Euroopan neuvoston hyväksyttäväksi Rooman sopimukseen vuodelta 1957 perustuvan menettelytavan mukaisesti. Myöhemmin Maastrichtin sopimus 1986, Amsterdamin sopimus 1999 ja Lissabonin sopimus 2009 lisäsivät Euroopan parlamentin valtaa

lainsäätämisen prosessissa ja nykyisin Euroopan parlamentti ja Euroopan neuvosto hyväksyvät useiden alojen säädökset yhdessä. Komissio hoitaa säädösten täytäntöönpanon.³⁷ Asetukset hyväksytään suoraan osaksi jäsenvaltioiden kansallista lainsäädäntöä voimaantulopykälällä: ”Tämä asetus tulee voimaan kahdentenakymmenentenä päivänä sen jälkeen, kun se on julkaistu Euroopan unionin virallisessa lehdessä. Tämä asetus on kaikilta osiltaan velvoittava, ja sitä sovelletaan sellaisenaan kaikissa jäsenvaltioissa.” (Euroopan unionin virallinen lehti 4.12.2008, L 326/13). Asetukset eivät direktiivien tapaan käy läpi kielellistä muokkausta jäsenmaissa ennen niiden integrointia osaksi säädöskokoelmaa.

3.6 Yhteisesiintymien rekisterisidonnaisuus

Kääntäjän kannalta haasteellista yhteisesiintymissä on, että ne ovat usein hyvin rekisterisidonnaisia. Yhdessä erikoiskielessä taaajaan yhdessä esiintyvät elementit voivat toisessa tai yleiskielessä olla odotuksenvastaisia ja tuottaa tahattomia konnotaatioita (semanttisesta prosodiasta tarkemmin alaluvussa 2.2.3). Erikoiskielen ja sen konventioiden hyvä tuntemus tai laaja yhteisesiintymiä sisältävä käännösmuisti auttaa kääntäjää varmistamaan, että kohdekielinen ilmaus on idiomaattinen ja odotuksensemukainen. Kuten Baker huomauttaa, ei riitä, että kääntäjä hallitsee erikoiskielen tai ammattikielen termit, hänen tulee myös hahmottaa näiden termien yhteisesiintymien valikoima (collocational range). Bakerin esimerkki on peräisin englanninkielisestä tietotekniikan erikoiskielestä: kääntäjän tulee tietää, että englannin kielessä *data* on substantiivi, joka yhteisesiintyy verbien *handle*, *extract*, *process*, *manipulate* ja *retrieve* kanssa ja muodostaa yhdyssanoja substantiivien *bank* ja *processing* kanssa. (Baker (2011): 57). Termillä voi olla muulla elämänalueella tai toisessa erikoiskielessä eri merkitys, joten yhteisesiintymä on aina kontekstisidonnainen. Sama pätee myös oikeuskielen sisällä: termillä voi olla toisella oikeudenalalla eri merkitys kuin mikä jonkin alan tai yleisestä termisanakirjasta löytyvällä sanakirjaesiintymällä on (Mattila (2002): 542).

3.7 Käännökset tutkimuskohteena

Käännöskielen tutkimus nousi Suomessa kiinnostuksen kohteeksi 2000-luvun alussa Joensuun yliopiston tutkimushankkeen ”Käännössuomi ja kääntämisen universaalit – tutkimus korpusaineistolla” myötä. Hankkeen tavoitteena oli tutkia käännöstieteen tutkimuskentälle 1990-luvulla tuodun hypoteesin käännösuniversaaleja, kaikille käännösten tekstilajeille ja niitä edustaville teksteille yhteisiä piirteitä (Baker

³⁷<http://europa.eu>

(1993), Baker (1996), Laviosa (1996)). Tutkimusryhmän jäsenten tutkimuskohteet kattoivat syntaktisia ja leksikaalisia ilmiöitä ja ideologian tutkimusta useista eri lähtökielistä ja genreistä suomennetuissa teksteissä (Eskola (2002), Jantunen (2004), Kemppanen (2008)). Universaalien tutkimuksessa kysymystä samuudesta tai eroavaisuudesta voidaan lähestyä kahta kautta. Lähtökielisten tekstien ja niiden käännösten kontrastiivisen tutkimuksen tavoitteena on tunnistaa esimerkiksi interferenssiä, käännösten standardisoitumista tai eksplikoitumista, jotka ovat Chestermanin luokituksen mukaan S(ource language)-universaaleja. T(target language)-universaalit kattavat muun muassa kohdekielelle tyypillisten ominaispiirteiden aliedustuksen tai kohdekielelle epätyypilliset leksikaaliset rakenteet ja niitä tutkitaan vertailemalla käännöksiä alunperin kohdekielellä kirjoitettuihin teksteihin. Chesterman (2004). Tutkimuksessaan akateemisten tekstien käännöksistä verrattuna alunperin kohdekielellä kirjoitettujen saman tekstilajin teksteihin vuodelta 2000 Anna Mauranen esittää kaksi hypoteesia, jotka molemmat ovat tutkimukseni kannalta keskeisiä. Ensimmäisen hypoteesin mukaan käännökset sisältävät epätyypillisiä sanayhdistelmiä, kuten kollokaatioita, ja alustavat tulokset osoittavat niitä löytyvän aineistoista, jopa siinä määrin, että epätyypillisen aineksen esiintyminen käännöksissä voisi olla eräänlainen kääntämisen universaali ja pohtii tilastollisin perustein määräytyvän käännösuniversaalin potentiaalia. Toisen hypoteesin mukaan kääntäjät ovat paremmin selvillä siitä, mikä on käsillä olevan vastineen valinnassa kohdekielellä mahdollista kuin siitä, mikä on tyypillistä. Muut tutkimuksen havainnot koskivat kohdekielellä taajaan esiintyvien leksikaalisten elementtien yliedustusta ja suurempaa vaihtelua lähisyonymien käytössä käännöksissä kuin alkukielellä. Mauranen huomauttaa myös, että kontrastiivisen tutkimuksen vaarana on kiinnittää huomiota eroavaisuuksiin samuuden kustannuksella. (Mauranen (2000). Kollokaatiotutkimus sivuaa synonymian tutkimusta aina, kun tutkimus valottaa kollokaatioiden vaihtelua ja variaatiota. Jantusen tutkimus astemääritteitä *hyvin, kovin ja oikein* paljastaa kontekstuaalisia piirteitä, jotka erottavat synonymiset astemääritteet toisistaan (Jantunen (2004)). Käännökset paljastavat kielellisen merkityksenannon mekanismeja. Firthin puheen otsikko ”*Linguistics and Translation*”: ”translation as an activity revealing aspects in the nature of language” (Palmer (1968): 75) on ollut merkityksellinen tämän tutkimuksen käynnistäjänä. Yhdistettynä ajatukseen kollokaatiosta nimenomaan merkitystä luovana voimana otsikko herätti aikanaan kysymyksen siitä, miten asiaa voisi tutkia käännöksistä ja onko käännetyissä teksteissä kääntämisen universaalien aiheuttamaa samankaltaisuutta. Samaisessa artikkelissa Anna Mauranen esitti myös ajatuksen aineistojen suppeudesta ja tarpeesta käsitellä kysymystä laajemmilla tekstikorpuksilla (Mauranen (2000): 12). Kontrastiivisen kielimallin näkökulmasta tarvitaan paljon kontekstualisoitua tutkimusta, jonka tulokset eivät ole yleistettävissä koskemaan kieltä kyseistä tekstilajia laajemmin.

Firthin sanoin, ”The statement of meaning by interlinear word for word or literal

translation as an aid to analysis is equally inadmissible though often still met with.” (Palmer (1968): 75). Merkitys kätkeytyy yhtä sanaa tai merkkiä isompaan rakennelmaan, joten havainnot sanasanasaisesta käännöksestä paljastavat piirteitä kielen olemuksessa.

Kollokaatiotutkimus erikoiskielten kuvaamisen lähestymistapana soveltuu hyvin kääntämisen tutkimukseen (Palmer (1968): 109–113). Potentiaalisten merkitysten kenttä on erikoiskielissä pienempi ja paremmin hahmotettavissa kuin yleiskielessä. Uskon Firthin tapaan, että käännöksiä analysoimalla päästään lähelle kielen järjestyksen mekanismeja ja kieltenvälisiä eroja. Omassa tutkimuksessani keskityn kuitenkin arvioimaan asetusten suomennosten tai suomenkielisten versioiden samankaltaisuutta verrannaisaineiston sisältämien alunperin suomeksi kirjoitettujen säädösten kanssa. Tutkimusaineiston sisältämien suomennosten lähtökieli ei ole tiedossa, mutta tiedetään, että liittymisen aikaan käännöksiä tehtiin pääosin ranskasta, englannista ja saksasta, katso alaluku 1.4). Kieltenvälinen kontrastiivinen ote on siitä syystä tämän tutkimuksen ulkopuolella.

Tutkimukseni aineisto on syntynyt kääntämisen seurauksena, se on käännöskieltä. Euroopan yhteisöön (nyk. Euroopan unioni) liittyvän Suomen oli tuotettava nopeasti suuri määrä yhteisöainsäädännön suomennoksia ja EU:n asetukset integroitiin suoraan suomalaisen säädöstekstin lomaan täydentämään ja korvaamaan kansallista lainsäädäntöä. Tutkimusaineisto koostuu näistä asetuksista. Asetuksista tiedetään niiden aihealue silloisen jaottelun perusteella sekä asetuksen hyväksymisvuosi. Muutoin niiden kääntämisestä on saatavissa vain vähän tietoa, mikä on EU-kontekstissa tyypillistä. Kääntäjän rooli EU:ssa on toimia suhteellisen näkymättömänä välittäjänä (Koskinen (2000a): 89). Aineistot on kuvattu tarkemmin luvussa 4 ja tutkimusmenetelmät luvussa 5.

3.8 Säädöskielen ja EU-kielen tutkimusta Suomessa

Heikki E.S. Mattilan tutkimustyö oikeustieteen ja lingvistiikan rajapinnassa oikeuslingvistiikassa on osaltaan auttanut omankin tutkimukseni aiheen rajaamisessa. Oikeustieteen käsitteistöä laajasti kattavan oikeustietosanakirjan *Encyclopaedia Iuridica Fennica* valmistuminen tarjosi ensimmäisen mahdollisen lähteen tutkittavien termien valintaa varten. Oikeustietosanakirja perustuu kuitenkin Suomen lainsäädäntöön ja sen ilmestyessä Euroopan unionissa käytetty kieli oli vasta rantautumassa Suomeen.

Kotimaisten kielten keskuksen EU-kielenhuoltaja Aino Piehl haastatteli tutkimuksessaan vuonna 2000 julkishallinnon virkamiehiä, jotka pitivät virkkeen pituutta suurimpana EU-asetusten vaikeaselkoisuuden syynä. Informantit pitivät selkeyttä

ja ymmärrettävyyttä haittaavina tekijöinä myös outoja ilmauksia ja termejä, jotka viittaavat sellaisiin käsitteisiin, joilla näyttäisi olevan jo valmiiksi suomenkielinen vastine (Piehl (2006): 191). Suomalaisia säädoskielen ymmärrettävyyden ja selkeyden tutkimuksia, joissa tutkimuskohteina ovat muun muassa lauseiden ja virkkeiden pituus, lauseiden määrä virkkeissä sekä alisteisten lauseiden ja nominalisaatioiden määrät ovat Piehlin mukaan muun muassa Mäkitalon, Naskalin ja Niemikorven tutkimukset vuosilta 1970, 1992 ja 1991 sekä Virtaniemen tutkimus vuodelta 1989, jossa hän vertasi lainsäädäntöä uutisten ja proosan yleiskieleen (raportoitu teoksessa Piehl (2006): 184–185). Tutkimuksen kohteena ovat olleet myös Saksan ja Suomen maakaarten käsitteelliset erot (Kinnunen (2006)) sekä Suomen lainsäädännön ruotsinnosten revisiot (Nordman (2009)).

Lisensiaatintyössään vuodelta 2002 Johanna Ketola kartoittaa työläinsäädännön sanaston tekstien ominaispiirteitä rekisteri- ja tekstilajiteorian mukaisesti systeemisfunktionaalisen kielimallin pohjalta. Tekstit ovat hänen mukaansa abstrakteja, niissä käytetään värittömiä verbejä ja merkitys selvää verbin komplementista. (Ketola (2002): 154). Ketola toteaa tutkimuksessaan myös, että lakikieli muuttuu yhteiskunnan muuttumisen myötä, kun on tarpeen säätää yhä moninaisempia tilanteita koskevia lakeja ja kuvata entistä abstraktimpia yhteiskunnallisia ilmiöitä. Kielen muuttumisen lisäksi sen välittämän informaation määrä kasvaa. (Ketola (2002): 20–21). Lakiteksti on oma tekstilajinsa, ja ”lakitekstin ymmärtämisen vaikeus voi lisäksi johtua yksittäisistä merkityksistä”, eikä kielentutkija voi maallikkona olla varma siitä, että on tulkinut merkityksen oikein (Ketola (2002): 156–158).

Tässä luvussa olen tarkastellut tutkimuskohteena olevien tekstien tekstilajia, säädoskieltä, sen ominaispiirteitä alun perin suomeksi kirjoitettuna ja käännettyinä tekstinä. Erityisesti käännetty EU-lainsäädäntö on sekä syntyhistoriansa että aineiston velvoittavuuden näkökulmasta omintakeista. Tutkimusaineiston tallentaminen käännosurakan alkuvaiheessa sekä silloisen lainsäädännön joidenkin osa-alueiden muuntaminen tietokoneella käsiteltävään muotoon on tuottanut sekä kääntämisen että kielen muutoksen suhteen ainutlaatuista materiaalia. Seuraavassa luvussa kuvaan tutkimusaineiston ja sen neljä verrannaisaineistoa, perustelen näiden valinnat ja esittelen tutkittavan termin.

Luku 4

Tutkimusaineistot ja tutkimuskohde

Suomi liittyi Euroopan unionin jäseneksi vuonna 1995. Tutkimuksen varsinainen tutkimusaineisto (EU) koostuu Euroopan unionin suomenkielisistä säädösteksteistä, jotka on liittymisvaiheessa liitetty osaksi Suomen säädöskokoelmaa. Suuri määrä Euroopan unionin (silloisen Euroopan yhteisön) asiakirjoja suomennettiin ennen liittymistä, muun muassa osa tämän tutkimuksen kohteena olevista asetuksista. Tutkimusaineiston vanhin asetus on alun perin säädetty vuonna 1958 ja uusin asetus on vuodelta 1999. Tutkimuksen ensimmäinen verrannaisaineisto (SL) on otos Suomen säädöskokoelmasta. Toinen verrannaisaineisto on Kielipankissa kesällä 2019 julkaistu Semanttinen Finlex-hankkeen versio Eduskunnan alkuperäissäädöksiä suomeksi vuosilta 1734–2018, Korp-versio. Kolmas verrannaisaineisto on Kotimaisten kielten keskuksessa koostettu, Kielipankin kautta käytettävissä oleva tekstikorpus Lakeja ja direktiivejä, Kielipankki-versio. Sen tehtävänä oli alun perin toimia ikkuna uudempiin käännöksiin, mutta aineiston kattama ajanjakso eroaa varsinaisesta tutkimusaineistosta niin vähän, että tämä tavoite ei toteutunut. Aineiston rooli tutkimuksessa onkin tarjota mahdollisuus tarkastella varsinaisia tutkimusaineistoja laajempaa tekstiyhteyttä ja direktiivejä aineistona, joka on käynyt läpi tekstin revisioprosessin ennen integrointia säädöskokoelmaan. Neljännen verrannaisaineiston, 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä, avulla saadaan yleiskielinen näkymä tutkimuskohteeseen. Tutkimuksen pääpaino on kuitenkin varsinaisen tutkimusaineiston (EU) tarkastelussa ja tulosten vertaamisessa ensimmäiseen verrannaisaineistoon (SL).

Tässä luvussa esittelen tarkemmin varsinaisen tutkimusaineiston (EU) ja sen neljä verrannaisaineistoa alaluvuissa (4.1–4.5), yhteenvedon aineistoista alaluvussa (4.6) sekä niiden valintaan vaikuttaneet seikat ja tutkittavaksi valitut kielelliset elementit alaluvussa (4.7).

4.1 Tutkimusaineisto EU ja ensimmäinen verrannaisaineisto SL

Varsinainen tutkimusaineisto (Westerlund 1999) koostuu Euroopan yhteisöjen asetusten käännöksistä³⁸ suomeen. Asetukset in käännetty Suomen Euroopan yhteisöihin liittymisen aikoihin 1990-luvun puolivälin jälkeen. Asetukset on säädetty ajanjaksolla 1958–1999 ja ne olivat vapaasti saatavilla verkossa EU:n EUR-Lex-tietokannassa. Osa asetuksista (59 kpl) oli vuonna 1999 saatavissa Kauppakaari Oy:n julkaisemalla CD-levyllä, mutta niitä ei ollut luokiteltu Euroopan unionin oman luokittelujärjestelmän mukaisiin lainsäädännön aihealueisiin – eikä niitä myöskään sisällytetty varsinaiseen tutkimusaineistoon. Tuolloin oli ladattavissa vielä varsin vähän uuden lainsäädännön suomenkielisiä käännöksiä, joten aineisto on koostettu saatavilla olevasta aineistosta siten, että lainsäädännön aihealueista mahdollisimman moni olisi edustettuna. Muutamasta osa-alueesta ei vielä ollut tekstejä lainkaan julkisesti saatavilla tai vastinetta verrannaisaineistossa. Käytän tutkimusaineistosta lyhennettä EU.

Suomen vuoden 1998 säädöskokoelmasta koostettu vastaavia aihe-alueita sisältävä aineisto (Kauppakaari 1999) on koottu Kauppakaari Oy:n CD-levyltä. Euroopan yhteisön jo Suomen lainsäädäntöön implementoitu lainsäädäntö (direktiivit ja asetukset) on poistettu aineistosta. Jäljelle on kuitenkin jäänyt säädöksiä, joita ei ole voitu tunnistaa EU-teksteiksi, koska ne ovat peräisin muun muassa Euroopan neuvoston (EN) päätöksistä. Käytän ensimmäisestä verrannaisaineistosta lyhennettä SL.

Kaikki varsinaisen tutkimusaineiston (EU) tekstit ovat käännöksiä ja verrannaisaineiston (SL) tekstit alun perin suomeksi kirjoitettuja säädöksiä. EU-aineisto sisältää Euroopan parlamentin ja neuvoston asetuksia ja SL-aineisto sisältää lakeja. Suomessa tehdyn liittymisvaiheen käännöstyön jälkeen kääntäminen siirtyi EU:n käännöstoimen tehtäväksi. Sen jälkeen julkaistut Euroopan unionin säädöstekstit on tuotettu yhtäpitäviksi dokumenteiksi lähdemateriaalista, joka on useimmiten koottu useista eri lähdedokumenteista monikielisestä aineistosta.

Aineistoissa edustettuna ovat osa-alueet ja aihealueet ja niiden koko saneina löytyvät taulukosta 1. Taulukon ensimmäinen osa sisältää EU-aineiston aihealueet ja niiden sisältämien asetusten lukumäärät. Toisessa osassa aineistot suhteutetaan toisiinsa, koska EU-aineiston aihealueet jakautuvat teksteihin eri tavoin kuin SL-aineiston. Jos aihealueen kohdalla on merkintä ”sisältyy”, vastaavat kyseisen aihealueen EU-asetukset SL-aineistossa laajempaa aihealuetta. Esimerkiksi EU-asetusten aihealueen 05 (Työvoiman vapaa liikkuvuus ja sosiaalipolitiikka) sisältämää lain-

³⁸uropa.eu.int/eur-lex/fi

säädäntöä löytyy Suomen säädöskokoelman kolmen osa-alueen säädöksistä: Eläke-, perhe- ja sosiaali-oikeus, Sosiaali- ja terveydenhuolto sekä Työ- ja virkasuhde. Ensimmäinen verrannaisaineisto (SL) on noin 2,5 kertaa varsinaisen tutkimusaineiston (EU) kokoinen.

Sananmuototyyppi/sananmuotoesiintymä-suhdeluku (*type/token*) kertoo aineiston sisältämästä toistosta, toisin sanoen aineiston sanastollisesta tiiviyydestä. Tunnusluvut on laskettu kirjaimista koostuvien saneiden lukumäärästä ja ne on tässä pyöristetty lähimpään sataan tuhanteen esiintymään. Yhteisesiintymien positiokoh- taisia esiintymistajuuksia laskettaessa olen käyttänyt laskentayksikkönä sanetta, muun muassa välimerkit, lyhenteet ja symbolit.

Aineistojen suhdeluvut ovat

EU: 25000/1300000; suhdeluku type/token noin 0,019, käänteisluku token/type 52 (toiston määrä)

SL: 44000/3500000; suhdeluku noin 0,012, käänteisluku token/type 79 (toiston määrä)

EU-aineisto on noin 1,5 kertaa tiiviimpää kuin SL-aineisto ja SL-aineistossa on enemmän toistoa ($79/52=1,519$). Näistä sananmuototyypeistä noin 11 000 on yhteisiä elementtejä, joten EU-aineistossa on noin 14 000 sellaista sanaa, joita SL-aineistosta ei löydy (Westerlund (2006a)). Vaikka huomioidaan, että tekstit edustavat eri aineistoja, ovat ne kuitenkin saman säädöstyypin tekstejä. Voidaankin esittää tutkimushypoteesia tukeva oletus, että EU on tuonut Suomen säädöskieleen uusia käsitteitä.

4.2 Toinen verrannaisaineisto: semfinlex-lait

Kielipankin Semfinlex-aineisto **Eduskunnan alkuperäissäädöksiä suomeksi vuosilta 1734-2018, Korp-versio**³⁹ (Helsingin yliopisto, FIN-CLARIN (2019)) pohjautuu Semanttinen Finlex -hankkeen⁴⁰ julkaisemiin avoimen datan aineistoihin. Aineistoperhe sisältää eduskunnan alkuperäissäädöksiä, korkeimman oikeuden ja korkeimman hallinto-oikeuden päätöksiä suomeksi ja ruotsiksi eri ajanjaksoilta sekä alkuperäissäädösten suomi-ruotsi-rinnakkaiskorpuksen. Sen sisältämät aineistot on jaettu tekstilajin perusteella osakorpuksiin. Tutkimuksen toinen verrannaisaineisto koostuu alkuperäissäädöksiä sisältävän osakorpuksen teksteistä, joiden säädöstyypit on laki. Muut korpuksen tekstilajit ovat asetus, päätös, työjärjestys, ilmoitus, kaari, luettelo, kirje ja kuulutus. Asetukset on jätetty haun ulkopuolelle, koska ne voivat

³⁹<http://urn.fi/urn:nbn:fi:lb-2019042601>

⁴⁰<https://data.finlex.fi/fi/projekti>

Taulukko 1. Tutkimusaineiston (EU) aihealueet ja verrannaisaineiston (SL) aihealueet sekä molempien aineistojen saneiden lukumäärät.

EU aihealue nro	Aihealue ja asetusten lukumäärä	
01	Yleiset sekä varainhoitoa ja toimielimiä koskevat asiat (17)	
02	Tulliliitto ja tavaroiden vapaa liikkuvuus (38)	
03	Maatalous (150)	
04	Kalastus (31)	
05	Työvoiman vapaa liikkuvuus ja sosiaalipolitiikka (20)	
06	Sijoittautumisoikeus ja palvelujen tarjoamisen vapaus (9)	
07	Liikenne (42)	
08	Kilpailupolitiikka (18)	
09	Verotus (7)	
10	Talous- ja rahapolitiikka sekä pääoman vapaa liikkuvuus (21)	
11	Ulkosuhteet (0)	
12	Energia (6)	
13	Teollisuuspolitiikka ja sisämarkkinat (8)	
14	Aluepolitiikka ja rakenteellisten välineiden yhteensovittaminen (0)	
15	Ympäristö, kuluttajat ja terveydensuojelu (7)	
16	Tiede, informaatio, koulutus ja kulttuuri (4)	
17	Yritysoikeus (4)	
18	Yhteinen ulko- ja turvallisuuspolitiikka (2)	
19	Yhteistyö oikeus- ja sisäasioissa (1)	
20	Kansalaisten Eurooppa (0)	
21	Muut (luokittelemattomat) (59)	

EU aihealue	EU saneita	SL	SL saneita
05	11 4334	Eläke-, perhe- ja sosiaalioikeus	21 7745
01	49 201	Hallinto-oikeus	134 131
02	264 736	Kansainväliset suhteet	76 894
03	325434	Maatalous	214 662
04	42 893		
16	8 509	Opetus- ja sivistystoimi	199 216
10	51 829	Rahoitus	270 018
05	sisältyy	Sosiaali- ja terveydenhuolto	269 528
18	2 208	Turvallisuus ja yleinen järjestys	157 418
19	956		
05	sisältyy	Työ- ja virkasuhde	279 075
14	sisältyy	Valtio-oikeus	148 045
09	2 309	Verolaki	371 269
06	19 886	Viestintä, liikenne ja kuljetus	314 795
07	118 302		
01	sisältyy	Yleishallinto-oikeus	49 888
12	12 811	Ympäristö ja asuminen	449 939
15	24 163		
02	sisältyy	Yritystoiminta	409 369
08	88 626		
13	26 182		
17	46 786		
21	162 314		
	Saneita yht.		Saneita yht.
	1 371 479		3 561 992

mahdollisesti olla EU-taustaisia ja tarkoituksena on pysyä säädöstekstien tutkimuksen parissa tässä verrannaisaineistossa. Käytän kolmannesta verrannaisaineistosta lyhennettä semfinlex-lait.

4.3 Kolmas verrannaisaineisto: LaDi

Sisällytin kolmannen verrannaisaineiston **Lakeja ja direktiivejä, Kielipankki-versio**⁴¹ (Kotimaisten kielten tutkimuskeskus (2016)), koska osa sen sisältämisestä säädöksistä edustaa varsinaisen tutkimusaineiston jälkeisenä ajanjaksona Suomessa tuotettuja tekstejä. Aineisto koostuu laeista, asetuksista ja direktiiveistä vuosilta 2002–2003. Alkuperäisenä ajatuksena oli löytää esimerkkejä asetusten suuren käännösurakan jälkeiseltä ajalta vuodesta 2000 eteenpäin. Kielipankin kautta käytettävissä oleva Kotimaisten kielten keskuksen koostama tekstikorpus toteuttaakin osittain alkuperäisen tavoitteen, joskin aineisto sisältää myös liittymisvaihetta edeltävän ajanjakson tekstejä, kuten taulukosta 2 käy ilmi. Ennen vuotta 2000 tuotettuja Euroopan unionin asetuksia ei sisälly varsinaiseen tutkimusaineistoon (EU-asetuksia) eikä kyseisiä tekstilajeja sisälly lainkaan ensimmäiseen verrannaisaineistoon (SL). Käytän kolmannesta verrannaisaineistosta lyhennettä LaDi.

Käytän kolmannen verrannaisaineiston sisältämisestä Euroopan unionin tekstien, direktiivien, asetusten ja lakien luokista nimitystä **tekstilaji**⁴². Aineiston tekstilaji jakaantuvat alalajeihin siten, että direktiivien alalajeja ovat Euroopan parlamentin ja neuvoston direktiivi, Komission direktiivi ja Neuvoston direktiivi; asetusten alalajeja ovat Euroopan parlamentin ja neuvoston asetus ja Neuvoston asetus. LaDi-aineisto sisältää sekä käännöksiä (asetukset, direktiivit) että suomeksi kirjoitettuja säädöstekstejä (lait). LaDi-aineiston direktiivien ja asetusten sisältämät tutkimuskohteiden esiintymät on luokiteltu ja laskettu tekstilajeittain, jotta mahdolliset poikkeamat ensimmäisen analyysin tuloksista olisivat mahdollisimman selkeitä.

Taulukossa 2 esitetään tekstien lukumäärät tekstilajeittain kahdelta ajanjaksolta: vuosilta 1980–1999, joka on päällekkäinen ajanjakso varsinaisen tutkimusaineiston ja ensimmäisen verrannaisaineiston tekstien kanssa, ja vuosilta 2000–2002. Asetukset vuodelta 1999 eivät sisälly tutkimuksen varsinaiseen tutkimusaineistoon (EU). Myöskään Kotuksen tekstikorpuksen direktiivit eivät sisälly varsinaiseen tutkimusaineistoon, joka sisältää pelkästään asetuksia. Direktiivit on todennäköisesti myös käännetty osana liittymisvaiheen käännösurakkaa, mutta ne ovat käyneet Suo-

⁴¹<http://urn.fi/urn:nbn:fi:lb-20140730126>

⁴²Tekstilaji; Määritelmä 1) Rakennetta korostava määrittely: Saman tekstilajin tekstit koostuvat tietyistä vaiheista, joilla on vakiintuneita järjestyksiä ja jotka ovat eri tavoin pakollisia. Tiettyyn tekstilajiin kuuluvilla teksteillä on sama nimi, päämäärä/funktio, sama kohdeyleisö, prototyypillisesti samankaltainen rakenne ja sisältö. (Lähde: Tieteen termipankki 6.11.2016: Kielitiede:tekstilaji.)

Taulukko 2. LaDi-aineiston jakautuminen tekstilajeihin laki, asetus tai direktiivi sekä niiden julkaisemisen ajanjaksot ja aihealueet säädösten nimien perusteella.

Tekstilaji	Tekstit ajanjaksoista	Tekstien lukumäärä	Aihealueita
Aineiston koko (saneet)		267431	
Direktiivi	1980–1999	15	Maatalous. Työ- ja virkasuhde. Viestintä, liikenne, kuljetus. Yritystoiminta. Ympäristö ja asuminen.
Direktiivi	2000–2002	23	Kansainväliset suhteet. Maatalous. Rahoitus. Verolaki. Viestintä, liikenne, kuljetus. Yritystoiminta.
Direktiivit yhteensä		37	
Asetus	1999	2	Yleishallinto-oikeus.
Asetus	2000–2002	2	Ympäristö. Rahoitus.
Asetukset yhteensä		4	
Laki	2000–2003	25	Maatalous. Työ- ja virkasuhde. Verolaki. Viestintä, liikenne, kuljetus. Yleishallinto-oikeus. Ympäristö ja asuminen.

nessa läpi erillisen implementaatioprosessin, jossa ne on muutettu osaksi Suomen lainsäädäntöä, katso esimerkiksi Piehl (2013). Direktiivit ovat mukana aineistossa sen vuoksi, että niiden kautta saadaan kapea näkymä aineistoon, joka on Suomessa käynyt läpi lainsäädäntöprosessin. Katsauksen myötä saadaan hiukan tietoa siitä, onko direktiiveissä havaittavissa samankaltaisia yhteisesiintymiä kuin varsinaisessa tutkimusaineistossa ja muissa verrannaisaineistoissa.

4.4 Neljäs verrannaisaineisto: lehdet-osa

Neljännän verrannaisaineiston **1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä**⁴³ (Helsingin yliopisto (2017)) tehtävänä on tarjota tietoa tutkimuskohteista yleiskielessä. Kielipankin tekstikorpus **1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä, versio 2** sisältää kaksi kokoelmaa: tiedelehdet ja muut lehdet. Tutkimuksen verrannaisaineistoksi soveltuu muita lehtiä sisältävä kokoelma, koska tiedelehtien kielen voidaan olettaa edustavan kyseisten alojen erikoiskieltä. Muiden lehtien kokoelma sisältää yhteensä 270 lehteä. Niiden joukossa on myös Virallinen lehti, joka lehden sivuston sanoin ”huolehtii lakisääteisen ilmoitusvelvoitteen täyttämistä” (www.virallinenlehti.fi). Kielipankin korpustyökalu Korpin kautta käytettävissä olevassa korpuksessa voi valita listalta haun kohteeksi tulevat lehdet, jolloin on mahdollista jättää jokin lehti tai jotkin lehdet pois haun piiristä. Tässä tutkimuksessa Virallinen lehti ei ole haun kohteena olevien lehtien joukossa. Sen sijaan sitä on tarkasteltu erikseen lainalan erikoiskielen edustajana.

4.5 Aineistojen koko, edustavuus ja populaatio

Tässä osiossa pohdin tutkimusaineistojen edustavuutta arvioimalla niiden kokoa, miten tyypillisiä tekstit ovat säädöstekstien populaatiolle tai miten niiden voidaan olettaa tarjoavan käyttökelpoisen verrannaisaineiston tutkimusaineistolle. Tarkastelen lisäksi aineiston teknistä laatua.

Varsinaisen tutkimusaineiston (EU) koko on noin 1,3 miljoonaa sanetta ja ensimmäisen verrannaisaineiston (SL) koko noin 3,5 miljoonaa sanetta (taulukko 3). Molemmat ovat nykytutkimuksen näkökulmasta pieniä, kun käytettävissä on jo isoja aineistoja ja korpustyökalujen puolella enemmän laskentakapasiteettia. Erityisesti tilastolliseen evidenssiin perustuva tutkimus edellyttää kattavan aineiston tai riittävästi havaintoja, jotta tulokset olisivat yleistettävissä. Miljoonan sanan suuruinen aineisto voi kuitenkin olla riittävä, kun tutkitaan vähintään kohtuullisen usein esiintyviä sanoja (Rayson (2002):40). On kuitenkin vaikea tietää, mitä tarkoitetaan sanoilla ’usein’ tai ’kohtuullisen usein’. Luonnehdintojen merkitys muuttuu ajan mittaan. Oma tutkimukseni yhdistää määrällisen tarkastelun tulokset manuaaliseen yhteisesiintyjien todentamiseen, mikä ei mahdollista kovin suurten havaintomäärien tutkimista. Tutkimusaineistot edustavat kuitenkin erikoiskieltä ja sisältävät kattamansa ajanjakson aikana tuotetuista teksteistä suuren osan käytettävissä olleista säädösteksteistä. Tutkimustuloksia ei ole tarkoitus yleistää kattamaan säädöskielen lisäksi muita aihealueita.

⁴³<http://urn.fi/urn:nbn:fi:lb-2016011101>

Taulukko 3. Tutkimusaineiston ja ensimmäisen verrannaisaineiston koot, tiedostojen lukumäärät ja aineiston osa-alueiden lukumäärät.

	EU	SL
Aineiston koko (saneet)	1 371 479	3 561 992
Tiedostojen lukumäärä	444	15
Aineiston osa-alueiden lukumäärä	18 + muut	15

Varsinainen tutkimusaineisto (EU) jakautuu epätasaisesti EU:n lainsäädännön aihealueisiin. Syynä tähän on aineiston koostaminen siinä vaiheessa, kun vasta osa käännettävästä materiaalista oli valmiina suomenkielisellä verkkosivustolla. Erityisen paljon suomennettuja asetuksia oli saatavissa maatalouden aihealueelta, kun taas esimerkiksi verolainsäädäntöä tai ympäristöön liittyviä asetuksia oli huomattavan vähän saatavissa. Asetustekstien lukumäärä ei myöskään korreloi kovin hyvin niiden sisältämien sanamäärien kanssa: yritysoikeuden neljä tiedostoa sisältävät suunnilleen yhtä paljon saneita kuin kalastuslainsäädäntö, joka koostuu 31 asetuksesta. Lyhyissä asetuksissa on myös suhteessa enemmän rutiininomaisia vakiolausekkeita ja metatietoa. Eepätasaisuuden syynä ovat alkuaikojen käänntöyön laajuus ja käänntösten aihealueiden priorisointi. Aineisto on pyritty painottamaan aihealueittain siten, että tutkimusaineistoon on valittu asetuksia samassa suhteessa kuin niitä on ollut kyseisen aihealueen piiristä saatavilla aineiston kokoamisen aikaan.

Varsinainen tutkimusaineisto (EU) sisältää noin 80 prosenttia vuonna 1999 saatavilla olleista suomennetuista asetuksista. Kumpikaan korpuksista ei enää millään mittarilla mitattuna edusta lainsäädännön koko populaatiota lainsäädännön koko ajan lisääntyessä ja uusien asetusten korvautessa entisiä. Tutkimuskysymyksen valiteetti ei kuitenkaan tässä tapauksessa riipu aineiston ajantasaisuudesta, kun molemmat aineistot ovat samalta ajalta ja arvelen, ettei tutkittava ilmiö myöskään ole erityisen herkkä ajan tuomille muutoksille. Kieliteknologisten menetelmien kehittymisen ja aineistojen lisääntyvän saatavuuden myötä tätä kysymystä olisi syytä tarkastella erikseen.

Ensimmäisen verrannaisaineiston (SL) populaatio on osa Suomen lainsäädäntöä vuodelta 1999. Verrannaisaineisto tarkoittaa korpusta, jonka voidaan katsoa tarjoavan kielenkäyttäjälle ja -tutkijalle normeja vertailua varten (reference corpus) (Sinclair (1996)). Tästä aineistosta on kopioinnin jälkeen poistettu osiot, joiden aihepiiriin liittyvää lainsäädäntöä ei Euroopan yhteisöissä tuolloin ollut saatavilla EuroLexissä tai joita oli yksittäisiä asetuksia. Näitä osioita ovat muun muassa prosessioi-

Taulukko 4. Toisen verrannaisaineiston (semfinlex-lait) koko, tekstien lukumäärät ja aineiston osa-alueiden lukumäärät.

Ajanjakso	Aineiston koko	Lakien osuus
1960–1998	9 304 874	3 822 946
1999–2009	8 318 305	3 690 320
2010–2018	7 743 051	4 010 329
Yhteensä	25 366 230	11 523 595

keus ja siviilioikeus sekä direktiivit. Aineisto ei sisällä lain valmisteluaineistoja eikä myöskään muun muassa tuomioistuinten päätöslausemia tai lain selityskirjallisuutta. SL-aineistoa on tutkittu EU-aineiston kanssa rinnakkain ja aineistoja on käsitelty samalla tavoin.

Suomen lainsäädäntöä on itsenäistymisen ajoista asti kirjoitettu suomeksi, vaikkakin suuri osa alkuperäistä lainsäädäntöämme on aikoinaan käännetty ruotsin kielestä suomeen. Lainsäädännössämme on siten jo alun perin ollut käännettyä aineistoa. Ruotsin kielen vaikutus Suomen oikeuskieleen ei kuitenkaan uskoakseni ole sen suurempi kuin ruotsin vaikutus suomen kieleen ylipäätään, jolloin Suomen säädöskokoelmaa voi pitää alun perin suomeksi kirjoitettuna aineistona. Verrannaisaineistossa ei ole ruotsista suomeen käännettyjä säädöstekstejä. Verrannaisaineiston eli referenssikorpuksen tulee olla riittävän iso ja sen tulee tarjota kattavaa tietoa tutkimuskohteesta kohteena olevassa populaatiossa. Säädoskielen populaatio muodostuu säädöskokoelmaan kirjatusta säädöksistä, lainsäädännöstä. Koska Suomen säädöskokoelma kattaa koko virallisen lainsäädännön niiltä aihe-alueilta, joita tutkimusaineisto sisältää, voidaan sitä pitää edustavana referenssikorpuksena.

Toisen verrannaisaineiston (Kielipankin Semfinlex-aineiston osakorpus, tässä lyhytnimellä semifinlex-lait) koko aineiston saneiden lukumäärä on 28 240 594. Tutkimusta varten on kuitenkin valittu ajanjaksot, jotka eivät kata koko aineistoa. Aineiston suurimman tekstilajin, säädöstyypin laki, teksteissä on yhteensä 13 911 433 sanetta 2 306 348 virkkeessä, joten se muodostaa tällä hetkellä laajimman käytettävissä olevan tutkimuksen edustamaa tekstilajia sisältävän aineiston. Se tarjoaa näin ajantasaisen verrannaiskorpuksen varsinaisesta tutkimusaineistosta (EU) ja ensimmäisestä verrannaisaineistosta (SL) saaduille tutkimustuloksille. Kielipankin konkordanssityökalu Korp mahdollistaa hakujen rajaamisen ajallisesti halutulla tavalla, joten tutkimuskohteiden haut on toteutettu kolmen eri ajanjakson osalta erikseen. Nämä on taulukoitu taulukkoon 4. Ensimmäinen verrannaisaineisto (SL) on kooltaan noin seitsemäsosa toisesta verrannaisaineistosta (semfinlex-lait).

Taulukko 5. Kolmannen verrannaisaineiston (LaDi) koko, tekstien lukumäärät ja aineiston osa-alueiden lukumäärät. Aineiston koko on 267 431 sanetta.

	Direktiivit	Asetukset	Lait
Tekstien lukumäärä	37	4	25
Aineiston osa-alueiden lukumäärä	8	6	6

Kolmas verrannaisaineisto (LaDi) on Kotuksessa kerätty pieni, mutta hyvin dokumentoitu ja teknisesti laadukas aineisto, joka on käytettävissä Kielipankissa Korp-konkordanssityökalun kautta. Aineisto sisältää neljä Euroopan yhteisöjen asetusta ja 25 lakia kuudelta lainsäädännön osa-alueelta. Direktiivejä on enemmän kuin asetuksia ja lakeja yhteensä (ks. taulukko 5). Ensimmäisen verrannaisaineisto (SL) on kooltaan yli 15-kertainen kolmanteen verrannaisaineistoon (LaDi) verrattuna.

Neljäs verrannaisaineisto, 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä on julkaistu Kielipankissa vuonna 2017 ja se sisältää tiedelehtiä ja muita lehtiä. Näistä tutkimuskohteiksi valittiin muut lehdet, koska tavoitteena oli tarkastella tutkimuskohteita mahdollisimman yleiskielisessä aineistossa. Muita lehtiä on yhteensä 270 ja saneita niissä on 167 289 418 (mukaan lukien Virallinen lehti). Vanhin numero aineistossa on vuodelta 1990 ja uusin vuodelta 2017. Pääosa aineistosta on 2000-luvulta.

4.6 Yhteenvedo: tutkimusaineistot

Tutkittavat tekstilajit ovat EU-asetukset (parlamentin ja neuvoston), joita on varsinaisessa tutkimusaineistossa (EU) ja kolmannessa verrannaisaineistossa (LaDi). Niistä tehtyjä havaintoja verrataan ensimmäiseen, toiseen ja kolmanteen verrannaisaineistoon (SL, LaDi ja semfinlex-lait), jotka kaikki sisältävät lakeja. Laki on Suomen säädöskokoelman yleisin säädös. Kolmas verrannaisaineisto (LaDi) sisältää asetusten ja lakien ohella direktiivejä. Neljäs verrannaisaineisto (lehdet-osa) sisältää mediatekstejä. Tutkimusaineistojen yhteenvedo löytyy taulukosta 6.

Tutkimuksen kestäessä aineistojen saatavuudessa on tapahtunut paljon kehitystä. Varsinainen tutkimusaineisto ja sen ensimmäinen verrannaisaineisto voidaan nykyisin nähdä lähes historiallisina korpuksina, ne kuvaavat vuosituuhannen vaihteen tilannetta. Tutkimustulokset ovat kuitenkin ajasta riippumattomia, ja nyt Kielipankin ansiosta käytettävissä olevilla aineistoilla voidaan peilata tilannetta ajallisesti suhteellisen pian tutkimusaineiston kokoamisen jälkeen (LaDi) ja laajaan kaikki al-

Taulukko 6. Tutkimusaineistot.

Aineisto	Ajanjakso	Saneita	Tekstityyppi
EU	1998	1 371 479	tutkimusaineisto: asetuksia
SL	1998	3 561 992	1. verrannaisaineisto: lakeja
semfinlex-lait	1960–2018	11 523 595	2. verrannaisaineisto: lakeja
- osa-alue	1960–1998	3 822 946	2. verrannaisaineisto: lakeja
- osa-alue	1999–2009	3 690 320	2. verrannaisaineisto: lakeja
- osa-alue	2010–2018	4 010 329	2. verrannaisaineisto: lakeja
LaDi	2010–2018	267 431	3. verrannaisaineisto: lakeja, direktiivejä ja asetuksia
lehdet-osa	1990–2017	167 289 418	4. verrannaisaineisto: aikakaus- ja sanomalehtiä

kuperäiset säädökset sisältävään aineistoon (semfinlex-lait). Koska kysymys säädöskielen ja yleiskielen suhteesta on kiinnostava, on valittu verrannaisaineistoksi myös Kielipankin aikakauslehtiaineisto (lehdet-osa). Tässä tutkimuksessa pääpaino näiden aineistojenkohdalla on pidetty määrällisessä tarkastelussa jo pelkästään aineistojen laajuuden ja tutkittavana olevan termin yleisyyden vuoksi. Aineistojen saatavuuden lisäksi myös tutkimusmenetelmien valikoima on ajan mittaan laajentunut. Tämän kehityksen ja valitut tutkimusmenetelmät kuvaan seuraavassa luvussa 5. Tämä luku jatkuu tutkimuskohteiden kuvauksella.

4.7 Tutkittava termi: *toimenpide*

Substantiivi *toimenpide* valikoitui tutkittavaksi hakusanaksi ensin WordSmithTools-ohjelman tuottamalta avainsanalistalta ja myöhemmin tilastollisia tunnuslukuja laskevan ohjelmiston BSP (Bigram Statistical Package) tuottamalta tilastollisesti merkitsevien yhteisesiintymien listalta (ks. tarkemmin menetelmäluke 5). Valintaperusteena oli termin esiintyminen tutkimusaineistossa riittävän taajaan sekä sen löytyminen molemmilta yllä mainitulta listalta. Tutkimuskohteen ja tutkimusverbien muodostamien yhteisesiintymien laadullinen analyysi käsityönä näytti myös olevan mahdollisuuksien rajoissa, kun esiintymien lukumäärät liikkuvat sadoissa tai korkeintaan tuhansissa. Niitä ei kuitenkaan ole kymmeniätuhansia, toisin kuin taajim-

min esiintyvää termiä *asetus*, jonka esiintymiä löytyy EU-aineistosta lähes 22 000 kappaletta.

Firth ehdotti huomion kiinnittämistä yhdyssanoihin ja fraaseihin, joissa esiintyy 'tavallisia sanoja', toisin sanoen "collocational compounds and collocational phrases in which common words appear" (Palmer (1968): 107). Tavallisuudella tulkitseen tässä tutkimuksessa Firthin tarkoittavan yleiskielen sanastoa kontrastina erikoiskielten terminologialle. Erikoiskielissä tai osakielissä tavalliset sanat saavat usein erityisen alakohtaisen merkityksen.

4.8 Tutkittavat yhteisesiintyjät: *ryhtyä, soveltaa, suorittaa, säätää, tehdä ja toteuttaa*

BSP-ohjelmiston tuottaman tilastollisen analyysin (esiintymien kynnysarvo 3) tuloksena aineistosta erottuu viisi hakusanan kanssa tutkimuksen menetelmän näkökulmasta riittävän usein esiintyvää verbiä, jotka valikoituvat tutkimuskohteeksi sen vuoksi, että niiden merkityseroista yhdessä hakusanan *toimenpide* kanssa ei ensi kuulemalta muodostu selkeää kuvaa. Nämä viisi verbiä ovat: *ryhtyä, soveltaa, suorittaa, säätää* ja *toteuttaa*. Hakusana *toimenpide* esiintyy tuloslistalla joko ennen verbiä tai heti sen jälkeen. Verbi *tehdä* valikoitui tutkimuskohteeksi myös sen vuoksi, että se esiintyy tuloslistalla ainoastaan hakusanan *toimenpide* jälkeen, jolloin vaikutti motivoitulta tutkia tarkemmin sen esiintymiä tutkimusaineistossa (EU) ja verrannaisaineistossa (SL). Verbi *tehdä* ei myöskään esiinny EU-aineistossa BSP:n tuottaman tilastollisen analyysin perusteella usein. Verbi *säätää* puolestaan on mielenkiintoinen siinä mielessä, että *toimenpide* ei ole säädös – eikä niitä siten oletettavasti säädetä. Sen sijaan säädetään lakeja tai asetuksia, jotka koskevat *toimenpiteitä*. Verbit *suorittaa* ja *toteuttaa* vaikuttavat synonyymeiltä tai lähisynonyymeiltä ja verbi *ryhtyä* saattaisi viitata toiminnan aloittamiseen. Verbit on kuvattu tarkemmin alaluvussa 4.8 ja taulukkoon 7 on koottu verbit tai vastaavat verbikantaiset substantiivit sekä tieto siitä, minkä hakusanan yhteydessä ne listalla esiintyvät. Taulukko sisältää termin *toimenpide* lisäksi myös neljän muun lekseemin tiedot, jotka tutkimuksen eri vaiheissa olivat mahdollisia tutkimuskohteita *toimenpiteen* lisäksi. Aineisto mahdollistaa jatkossa myös näiden termien välisen vertailun kollokaattianalyysin avulla.

Verbin *säätää* valikoituminen oli harkitsematonta ja kuvastaa tutkimuksen alkuvaiheen päätöstä olla tarkastelematta tutkimuksen kohteena olevia tekstejä ennen tutkittavien lekseemien valintaa. Valintatilanteessa jäi siten havaitsematta, että verbi *säätää* on eräs säädöskielen tavallisimmista verbeistä ja että se esiintyy usein tutkimuskohteen lähiympäristössä, mutta ei muodosta sen kanssa tutkimuskohteena olevaa yhteisesiintymää. Puhdas distributionaalinen tarkastelu ilman manuaalista

karsintaa ei siten osoittautunut mahdolliseksi erityisesti verbin *säätää* tarkastelussa (ks. luku 5).

Taulukko 7. Vertailu verbien *ryhtyä*, *soveltaa*, *suorittaa*, *tehdä* ja *toteuttaa* esiintymistä BSP:n tilastollisesti merkitsevien kielellisten yksiköiden yhteisesiintymien listalla hakusanojen *toimenpide*, *toimi*, *toiminta* ja *toiminto* yhteydessä. Verbi sijaitsee EU-aineistossa ennen hakusanaa (EU L) tai sen jälkeen (EU R). (lähde: BST BIGRAM FREQ3: TOIMI, TOIMINTA, TOIMINTO, TOIMENPIDE)

VERBI tai sen JOHDANNAINEN	EU L	EU R
RYHTYÄ	TOIMI TOIMENPIDE	TOIMENPIDE
RYHTYMINEN	-	-
SOVELTAA	TOIMENPIDE	TOIMENPIDE
SOVELTAMINEN	-	TOIMENPIDE
SUORITTAÄ	TOIMI TOIMINTA TOIMINTO TOIMENPIDE	TOIMI TOIMINTO TOIMENPIDE
SUORITTAMINEN	-	TOIMI TOIMINTO
TEHDÄ	-	TOIMENPIDE
TEKEMINEN	?	?
TOTEUTTAA	TOIMI TOIMINTA TOIMENPIDE	TOIMI TOIMINTA TOIMINTO TOIMENPIDE
TOTEUTTAMINEN	-	TOIMI TOIMENPIDE

Seuraavassa selvitän tutkimusverbien perusmerkitystä lyhyesti perustuen Kotimaisten kielten tutkimuskeskuksen Kielitoimiston sanakirjaan ⁴⁴. Sen perusteella synonyymisiä tai lähisyronyymisiä verbejä ovat *tehdä*, *soveltaa*, *suorittaa* ja *toteuttaa* kun taas verbeillä *ryhtyä* ja *säätää* on näistä poikkeava perusmerkitys: *ryhtyä* viittaa toiminnan aloittamiseen ja *säätää* lain voimaan saattamiseen.

ryhtyä: 1. työn, toiminnan aloittamisesta: ruveta.

soveltaa: sovittaa käytäntöön, käyttää (muuntaen) hyväksi, toteuttaa käytännössä.

⁴⁴<https://www.kielitoimistonsanakirja.fi/>

suorittaa: 1. tehdä, toimittaa; us. paremmin toisin.

säätää: 1. menettelystä jolla laki laaditaan ja saatetaan voimaan; 2. määrätä säädöksellä.

tehdä: 1. valmistaa, tuottaa, laatia, saada aikaan; 2. suorittaa, toteuttaa, panna toimeen.

toteuttaa: 1. saattaa toteutumaan.

Kielitoimiston sanakirjan määritelmien joukossa verbi **säätää** poikkeaa muista, sillä sen konteksti rajataan määritelmässä lain tai säädöksen piiriin. Verbien **suorittaa**, **tehdä** ja **toteuttaa** määritelmissä viitataan ristiin toisiinsa, jolloin kyseisiä verbejä voidaan pitää määritelmien perusteella synonyymisinä. Toisaalta myös sekä verbien **soveltaa** ja verbin **tehdä** synonyymiksi mainitaan verbi **toteuttaa**. Verbit **ryhtyä** ja **toteuttaa** korostavat toiminnan aloittamista.

Tutkimuksen tavoitteena on verbien valinnan kautta tarkastella yhteisesiintymän muodostumista erikoiskielessä. Verbien valinta tutkimuskohteeksi on tehty jo tutkimuksen alkuvaiheessa motivoimatta valintaa muulla tavoin kuin valitsemalla tulostiedoston riveiltä löytyviä sellaisia verbejä, jotka intuitiivisesti vaikuttivat olevan merkitykseltään tarpeeksi lähellä toisiaan ja jotka esiintyvät riittävän usein tutkitavan substantiivin tekstiyhteydessä.

Valitut verbit ovat *ryhtyä*, *tehdä*, *soveltaa*, *suorittaa* ja *toteuttaa*. Vaikka verbi *ryhtyä* myös intuition perusteella poikkeaa toisista verbeistä perusmerkitykseltään, katsoin tutkittavia verbejä valitessani sen olevan oman kielitajuni perusteella oikeuskielessä lähellä muita tutkittavia verbejä. Lisäksi verbi *säätää* tuli valituksi sen vuoksi, että sen esiintymistäajuus osana yhteisesiintymää (silloin käytössä olevan menetelmän tuloksissa) vaikutti erikoisen suurelta.

Seuraavassa luvussa tarkastelen tutkimuksessa käytettyjä tutkimusmenetelmiä. Kuvaan tutkimusmenetelmän valinnan prosessia ja prosessin aikana tapahtuneita muutoksia käytettävissä olevien aineistojen ja korpuslingvististen menetelmien lisääntyessä ajan mittaan.

Luku 5

Tutkimusmenetelmät

Yhteisesiintymisen tutkimukseen liittyy käsitys kielestä tilastollisesti mallinnettavissa olevana ilmiönä. Ilman tilastollista tietoa kiinnostuksen kohteena olevasta ilmiöstä voidaan todeta, että kielellinen rakenne tai esiintymä on kielessä mahdollinen ja ainakin tutkijan tai kääntäjän kielitajun mukaisesti edustaa kielen konventioita vaikkapa sananmuodostuksen tai jonkin kielen rakenteen osalta. Kääntäjän tekemiä valintoja on tutkittu kartoittamalla kääntäjän toimintaa käännöstyön aikana, esimerkiksi pyytämällä häntä ajattelemaan ääneen työskennellessään (Think Aloud Protocol -menetelmästä katso Kussmaul ja Tirkkonen-Condit (1995), Jääskeläinen (1999)). Käännöstekstien tutkimuksen on katsottu keskittyvän menneeseen aikaan ja virheiden ja interferenssin etsimiseen teksteistä, jolloin kohdetekstien tutkimus ei suoraan paljasta kääntäjän ajatuksenkulkua tai kerro käännösprosessista (Kussmaul ja Tirkkonen-Condit (1995): 177–178). Virheanalyysin sijaan tutkimuksen lähtökohtana on ajatus siitä, että kääntäjällä on aina käytettävissään vaihtoehtoisia käännösstrategioita ja ratkaisuja vastineiden valintaan, ja oleellista on silloin käytettävissä olevan tiedon laatu ja mahdollisuus arvioida tätä tietoa käsillä olevan käännöstehtävän toteuttamiseksi parhaalla mahdollisella tavalla.

Tässä luvussa kuvaan tutkimuksessa käyttämäni menetelmät ja työkalut. Olen edellä tarkastellut tekstilajia luvussa 3. Tutkimusaineisto ja verrannaisaineistot on kuvattu luvussa 4 ja hypoteesit ja tutkimuskysymykset johdantoluvussa 1. Aluksi tarkastelen korpuslingvistiikkaa menetelmänä. Sen jälkeen kuvaan tekstikorpusten koostamisen ja käsittelyn menetelmät.

5.1 Korpuslingvistiikasta kielentutkimuksen menetelmänä

Korpuslingvistiikka voi tarkoittaa kielentutkimuksen metodia, teoriaa tai jopa tieteenalaa (kriittisestä suhtautumisesta viimeksimainittuun, ks. Schönefeld (2011)). Tässä tutkimuksessa korpuslingvistiikka tulkitaan kieliaineistojen käsittelyn menetelmäksi, jonka avulla testataan tutkimushypoteesien validiutta kieliaineistoista tai muodostetaan uusia tutkimusolehtamuksia (Tognini-Bonelli (2001), Bowker ja Pearson (2002)). Korpuslingvistiikka on tuonut myös käänntieteelliseen tutkimukseen todellisen kielenkäytön (attested language) tutkimuksen deskriptiivisen tutkimusotteen työkaluksi. Käänntösten ominaispiirteitä voidaan korpuslingvististen menetelmien avulla havaita, selittää ja ennustaa. (Baker (1993), Saldanha (2009))

Kieliaineistoja kieliteknologisin työkaluin hyödyntävä tutkimus voi olla luonteeltaan korpuspohjaista, korpuslähtöistä tai aineistolähtöistä (Rayson (2002)). Korpuspohjainen (corpus based) tutkimus pohjautuu valmiiseen hypoteesiin, ja aineistosta tutkittava kielellinen esiintymä on etukäteen tiedossa. Korpuslähtöisessä (corpus driven) tutkimuksessa tavoitteena on käsittelemättömän aineiston pohjalta rakentaa uusia kielimalleja korvaamaan tai täydentämään olemassa olevia. (Rayson (2002): 15). Myös ennalta annettuihin kielitieteellisiin käsitteisiin, esimerkiksi synonymiaan, suhtaudutaan usein varauksella. Aineistolähtöisessä (data based) tutkimuksessa hypoteesit nousevat aineistosta, mutta jäsenysmenetelmiä hyödynnetään, milloin niistä on apua aineiston tulkinnaassa (Rayson (2002): 14–17). Kollokaatiotutkimukseni on edennyt sillä tavoin syklistesti, että tutkimuskohteen ja tutkittavien kollokaattiehdokkaiden valinta nousi aineistosta käsin korpuslähtöisesti. Tutkimusaihekin syntyi aivan toisen aineiston pohjalta EU-tekstien kohtaaman kritiikin kautta, mutta tutkimusmenetelmän asettamien rajoitusten vuoksi tutkimus on lähentynyt korpuspohjaista linjaa. Koen kysymyksen tärkeänä, sillä se heijastaa korpustutkimuksen ytimessä olevaa todellisuutta oma tutkimukseni osalta, jossa kieliteknologiset menetelmät nostavat aineistosta esiin esiintymiä konkordanssien muodossa, joiden sisältämä kieliaines on ainoa näkemäni osa korpuksesta. Huomio kiinnittyy eroavaisuuksiin yhteisen aineksen sijaan.

Korpus kuvaa autenttista tilannetta kielestä tiettyjen rajojen sisällä. Korpusaineisto on oma kontekstinsa, jossa konteksti määrittää merkitystä ja mahdollistaa kielen tarkastelun juuri tässä käyttökontekstissa, tekstiyhteydessä ja tilannekontekstissa. Korpuslähtöinen tutkimusote tarkoittaa käytännössä sitä, että haku- tai analyysiohjelmat tuottavat kielentutkijalle aineistosta näytteen analysoitavaksi. Tilastollisten menetelmien käyttökelpoisuuteen suhteessa aineistoon vaikuttavat tutkittava kieli ominaispiirteineen, merkkijärjestelmiin liittyvät kysymykset, aineiston esikäsittely (perusmuotoistaminen, jäsenys tai muu annotointi) ja aineiston tek-

ninen laatu. Tilastollisia menetelmiä hyödyntävässä kielentutkimuksessa havainnot ja niihin perustuvat johtopäätökset rakentuvat aina sen varaan, mitä tietoa aineistosta saadaan. (Abney (1995): 19). Griesin mukaan tutkimuskirjallisuudesta löytyy yli 80 erilaista muuttujien välisiä suhteita (association measures) mittaavaa tilastollista menetelmää (Gries (2015): 56). Tilastollisten menetelmien soveltuvuus oman aineistoni analysointiin onkin osoittautunut yhdeksi tutkimukseni kipukynnyksistä.

Menetelmällisillä valinnoilla ja rajoituksilla voidaan vaikuttaa tulosjoukkoon. Määrittelemällä tutkimusikkunan koko rajoitetaan sen tekstimassan kokoa, jonka sisällä kielellisten esiintymien (esimerkiksi ortografisten sanojen) tulee esiintyä, jotta ne sisällytetään laskentaan. Iso ikkuna saattaa tuottaa vääriä kandidaatteja ja/tai ylivoimaisen työmäärän, ja toisaalta taas pieni ikkuna saattaa jättää liikaa löydöksiä tulosten ulkopuolelle.

Kielipankin aineistojen konkordanssityökalu Korpin konkordanssien saatavissa olevan tekstiyhteyden laajuus on tavallisesti virke. Varsinaisessa tutkimusaineistossani (EU) pisin virke on 443 saneen pituinen (se on EU-asetusten alussa oleva preambula eli yhdenlainen vakiolauseke), mutta toisesta verrannaisaineistosta (LaDi) löytyy EU-tekstien osiosta virke, jossa on peräti 994 sanetta. On siis otettava huomioon se mahdollisuus, että yhteisesiintymän osien välissä oleva tekstimassa voi säädöskiellessä olla niin pitkä, että etsitty yhteisesiintyjä ei löydy tutkimusikkunan sisältä.

5.2 Korpusten koostaminen, käsittely ja arkistointi

Tekstikorpusten koostamisen tekninen toteutus kuvataan tässä luvussa. Aineistojen koostamisen kriteerit, aineistojen koot ja erityispiirteet tekstilajin edustajana löytyvät luvusta 4.

Tämän tutkimuksen varsinaisen tutkimusaineiston (EU) ja sen ensimmäisen verrannaisaineiston (SL) käsittely tehtiin useassa vaiheessa: ensin 2000-luvun alussa, toisen kerran vuonna 2005, kolmannen kerran vuonna 2015 ja vielä neljännen kerran vuonna 2019.

Tekstikorpusten koostaminen ja arkistointi

Tässä alaluvussa esitän tiedot aineistojen koostamisesta ja saatavuudesta. Aineistot on kuvattu sisällöllisesti tarkemmin luvussa 4

Varsinainen tutkimusaineisto, EU-asetusten kokoelma, on kopioitu käsin Euroopan yhteisöjen verkkosivustolta (eu.org, haettu 1998). Asetukset kopioitiin yksitellen omiksi tiedostoikseen ja kopioinnissa tarkastettiin tekstin ensimmäisen ja viimeisen

virkkeen yhtäpitävyys jokaisen tekstin osalta ja kopioinnin onnistuminen tarkastettiin pistokokein tiedostossa muutamasta kohtaa. Varsinaisen tutkimusaineiston versioita on kolme: 1) alkuperäinen tekstimassa, 2) jäsennetty (pintamuoto + perusmuoto) versio ja 3) jäsennetyn version tiedostoista yhdistetty versio. Versiot on sijoitettu Helsingin yliopiston ylläpitämälle palvelimelle. Käytän aineistosta tässä tutkimuksessa lyhennettä 'EU'.

Ensimmäinen verrannaisaineisto, Suomen säädöskielen aineisto, oli vielä 1900-luvun lopussa Kauppakaari Oy:n yksinoikeuden piirissä aineiston formaatin vuoksi. Silloin tekstit olivat saatavissa painotuotteena ja CD:nä (Suomen Lakikokoelma). Vaikka lain sisältö ei ole tekijänoikeuden alaista, oli Kauppakaarella tekijänoikeus formaattiin. Vuonna 2000 tehty sopimus kattaa aineiston käytön tämän tutkimuksen tarpeisiin. EU-aineiston sisältöä vastaavat lainalat on kopioitu säädöskokoelman sisältävältä levytä kahdeksaan tiedostoon, joista jokainen sisältää yhden lain osaluheen, esimerkiksi Ym (ympäristölainsäädäntö). Tekstimassa on pyritty rakentamaan siten, että se vastaa sisällöltään mahdollisimman tarkasti varsinaista tutkimusaineistoa. SL-aineiston lähdemateriaali on edelleen tallessa Kauppakaaren CD-levyllä ja levytä kopioidut varsinaista tutkimusaineistoa vastaavat osat on sijoitettu Helsingin yliopiston ylläpitämälle palvelimelle. Käytän aineistosta tässä tutkimuksessa lyhennettä 'SL'.

Toinen verrannaisaineisto, Eduskunnan alkuperäissäädöksiä suomeksi vuosilta 1734-2018, Korp-versio⁴⁵ on Kielipankissa vuonna 2019 julkaistu Semanttinen Finlex-hankkeen⁴⁶ aineistojen pohjalta koostettu tekstikorpus. Aineistosta on Kielipankin Korp-konkordanssityökalun laajennetun haun työkalulla rajattu haun kohteeksi tietty ajanjakso ja säädöstyypit, jolloin varsinainen toinen verrannaisaineisto rajautuu näiden valintojen kautta useammaksi osakorpukseksi. Korpus on avoimesti käytettävissä Korpin kautta. Käytän aineistosta tässä tutkimuksessa lyhennettä 'semfinlex-lait'.

Kolmas verrannaisaineisto, Lakeja ja direktiivejä, Kielipankki-versio⁴⁷ on Kotimaisten kielten keskuksen koostama jäsennetty korpus, joka sisältää EU:n direktiivejä ja asetuksia ja Suomen säädöskokoelman lakeja. Se on sijoitettu Kielipankkiin⁴⁸ ja on siellä vapaasti käytettävissä. Käytän siitä tässä tutkimuksessa lyhennettä 'LaDi'.

Neljäs verrannaisaineisto, 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä -korpus, versio²⁴⁹, on FIN-CLARIN-hankkeessa koottu ja Kielipankissa julkaisu jäsennetty aineisto, joka on avoimesti käytettävissä Kielipankin Korp-

⁴⁵<http://urn.fi/urn:nbn:fi:lb-2019042601>

⁴⁶<https://data.finlex.fi/en/project>

⁴⁷<http://urn.fi/urn:nbn:fi:lb-20140730126>

⁴⁸kielipankki.fi

⁴⁹<http://urn.fi/urn:nbn:fi:lb-2017091901>

konkordanssipalvelun kautta. Korpuksen sisältämät muut kuin tiedelehdet muodostavat tässä tutkimuksessa käytetyn verrannaisaineiston. Käytän aineistosta lyhennettä 'lehdet-osa'.

Kevyt jäsenitys

Myös kääntäjän tarpeisiin soveltuvat avoimet UNIX-pohjaiset tekstikorpusten käsittelyn työkalut mahdollistavat erilaiset haut tekstimassasta. Jo 1990-luvun lopussa oli lisäksi käytettävissä suomen kielen syntaktisia jäsentimiä. Sekä EU-aineisto että SL-aineisto jäsenettiin tuolloin Connexorin FDG-jäsentimen edeltäjällä Fin-Lite-jäsentimellä. (Tapanainen ja Järvinen (1997))

Jäsenintä tarvittiin aikanaan tuottamaan korpuksen saneista perusmuoto (lemma), esimerkiksi <toimenpiteen> toimenpide. Sanaluokkatietoa tai dependenssirelaatioita aineistosta ei tuotettu, vaan katsottiin perusmuotoistamisen riittävän tutkimuksen tarpeisiin. Saneet sijaitsevat kukin omalla rivillään tekstitiedostossa, katso esimerkki 1 alla, jossa esitetään ote EU-asetusten aineistosta ensimmäinen asetus numero 358R0001 perusmuotoistetusta aineistosta, rivit 69–91.

(1)

```
358R0001 <14.5.1998> 14.5.1998
<EUROOPAN> Eurooppa
<TALOUSYHTEISÖN> talousyhteisö
<NEUVOSTO> neuvosto
<,> ,
<joka> joka
<ottaa> ottaa
<huomioon> huomio
<perustamissopimuksen> perustamissopimus
<217> 217
<artiklan> artikla
<,> ,
<jonka> joka
<mukaan> mukaan
<neuvosto> neuvosto
<antaa> antaa
<yksimielisesti> yksimielisesti
<säännöt> sääntö
<yhteisön> yhteisö
<toimielimissä> toimielin
<käytettävistä> käyttää
<kielistä> kieli
<,> ,
```

Suomen perusmuotoistaminen oli jo 1990-luvun loppupuolella mahdollista tuottaa luotettavasti, minkä ansiosta virheiden määrä varsinaisessa tutkimusaineistossa ja ensimmäisessä verrannaisaineistossa jää suhteellisen vähäiseksi. Jäsennys tuotti kuitenkin jonkin verran virheellisiä perusmuotoja. Virheiden osuus on tarkastettu satunnaisotannalla ja se on noin kaksi prosenttia. Havaitut virheet johtuvat pintamuodon monitulkintaisuudesta, joten niiden osuus ei vaikuta tuloksiin. Konkordanssi hakee halutun perusmuotoisen kielellisen elementin määrittelystä tutkimuskunasta tai virkkeestä. Haun tulokset tulostetaan pintamuotoisena ja tarkistetaan manuaalisesti.

Jäsennyksen tuloksena olevissa tekstitiedostoissa sekä pintamuotoinen että perusmuotoinen teksti sijaitsevat samassa tiedostossa esimerkin 1 mukaisesti. Kaikki haut on tehty samasta tekstiaineistosta.

5.3 Tutkimusmenetelmän valinta

Tutkimusmenetelmän valinta perustuu lähtöoletukseen, jonka mukaan tutkimuksen kohteena olevan kielellisen yksikön ominaispiirteitä voidaan kartoittaa tutkimalla tarkemmin sen esiintymis- tai käyttöympäristöä, tekstiyhteyttä.

Kollokaation tunnistamiseksi voidaan valita jokin monista tilastollisen assosiaation mittaamiseen ja todentamiseen perustuvista menetelmistä tai yhteisesiintyjää voidaan järjestää erilaisiin luokkiin arvioimatta yhteisesiintymän tilastollista merkittävyyttä. Menetelmän valinta on tapauskohtaista ja riippuu monesta tekijästä, joita ovat Seretanin (Seretan (2011)) kokoaman katsauksen mukaan ainakin aineiston koko ja käsittely, havaitut ja odotetut frekvenssit, frekvenssien raja-arvot, kertaesiintymät (hapax legomena) ja haluttujen kollokaatioiden tyyppi. Yhtä kaikkiin tarkoituksiin sopivaa tilastollista menetelmää ei ole käytettävissä (Seretan (2011): 43). Antti Arppe esimerkiksi päätyi osin metodipainotteisessa tutkimuksessaan käyttämään useita tilastollisia menetelmiä hyödyntävää tilastollista monimuuttuja-analyysiä, tutkimuskohteenaan neljä ajatteluverbiä ja niiden käyttöyhteyksiä (Arppe (2008)).

Tutkittavien yksiköiden valinta on tässä tutkimuksessa luonteeltaan kvantitatiivista: tarkastellaan havaintoja, jotka nousevat merkitseviksi useamman tilastollisen menetelmän antamissa tuloksissa. Valintaan vaikuttavia tekijöitä ovat hakusanan ja sen yhteisesiintymän esiintyminen yhdessä niin usein, että tuloksissa on analysoitavaa, mutta toisaalta ei niin usein, ettei tulosjoukkoa kykenisi käsittelemään manuaalisesti. Havainnot validoidaan manuaalisesti. Yksittäiset havainnot nousevat näkyviin aineistossa tuloslistojen läpikäynnin yhteydessä. Myös ne voivat olla

kiinnostavia, koska ne osoittavat poikkeamia tendensseistä tai profileista tai tavasta tehdä asioita. Näin ne voivat paljastaa jotain muuta tilannekontekstiin liittyvää, esimerkiksi käänösprosessista, käytetyistä kääntämisen apuvälineistä tai tekstin tai aihealueen kulttuurisidonnaisuudesta.

5.4 Tutkimusmenetelmä prosessina

Tutkin tutkimuskohteen *toimenpide* tekstiyhteyttä tuottamalla perusmuotoistetusta tekstistä konkordansseja, joihin ohjelma tulostaa tutkittavan yksikön lisäksi viisi ortografista sanetta sen molemmilta puolilta. Menetelmää kutsutaan kwic-analyysiksi (key word in context) tai konkordanssianalyysiksi.

Konkordanssien tuottamiseen voidaan hyödyntää ns. stop-listaa hylkäämään tuloksista merkkijonot, esimerkiksi välimerkit tai muut symbolit, joita tutkija ei katso tarpeelliseksi ottaa mukaan hakutuloksiin. En ole käyttänyt stop-listaa kahdesta syystä. Ensinnäkin perusmuotoistetusta tekstistä ajetut konkordanssit muodostuvat hankaliksi tulkita ilman lausekkeiden ja fraasien rajoja määritteleviä välimerkkejä. Välimerkkien, pykälämerkkien ja numeroiden avulla on helpompi erottaa tuloksista sellaiset rivit, joissa tutkittava yksikkö ja mahdollinen yhteisesiintyjä sijaitsevat eri lausekkeissa, lauseissa tai virkkeissä. Silloin yhteisesiintyjä ei ole tutkittavan yksikön aito yhteisesiintyjä vaan kuuluu toiseen yhdistelmään ja rivi on siksi poistettava hakutuloksista. Säädoskielen yksi ominaispiirteistä on sen lausekkeiden, lauseiden ja virkkeiden pituus ja pitkät upotukset, jolloin välimerkeillä on konkordanssien tulkinassa iso rooli. Toisekseen pykälämerkki ja muut symbolit, kuten lain lyhennyksenä käytetty iso L-kirjain, ovat osa laki- tai säädoskieltä ja siten sanojen lailla merkitystä kantavia yksiköitä. Liebscherin tutkimusten mukaan suljettujen luokkien sanoilla ei ole merkitystä korpuksia keskenään verrattaessa, koska niiden osuus aineistoissa on samaa suuruusluokkaa (Liebscher ja Groppe (2003): 12). Sisällytän laskettaviksi kielellisiksi elementeiksi tässä tutkimuksessa kaikki saneet, myös välimerkit ja symbolit.

Ensimmäinen vaihe tutkimuksessa oli tutkittavan aineksen valinta. Alustava ajatus saneen *toimenpide* kiinnostavuudesta syntyi WordSmithTools-ohjelman tuottamista kollokaatioanalyysistä. Myöhemmin BSP/NSP-ohjelmaperheen ohjelmilla tehdyt haut varmistivat valinnan. Näiden lisäksi potentiaalisina tutkimuskohteina listalla esiin nousivat *toimi*, *toiminta* ja *toiminto* sekä erityisesti yhdyssana-analyysien yhteydessä *hanke*, *järjestely*, *menetelmä* ja *menettely*. Tutkimuskohteiksi valikoituivat aluksi synonyymiset *toimenpide* ja *toimi*, joiden keskinäinen vertailu potentiaalisten yhteisesiintymien tarkastelun lisäksi muodosti aluksi tutkimuksen rungon. Myöhemmin verrannaisaineistojen ja niiden myötä manuaalisesti tehtävän validoinnin raskauden vuoksi toisen substantiivin analyysi oli jätettävä kesken.

Myös muut potentiaaliset tutkimuskohteet tarjoavat alustavien tulosten valossa hedelmällisiä yhteisesiintymisen tutkimuskohteita säädöskielessä, erityisesti yhdyssanojen tutkimuksen kentällä.

Toisessa vaiheessa nousi esiin kysymys, millä menetelmillä aineistosta saadaan esiin haluttu tieto hakukohteesta. Listaan halutun tutkimusikkunan sisällä esiintyvät saneet konkordansserilla ja etsin kahdesta tulostiedostosta uniikin ja toisaalta yhteisen aineksen vertailutyökalulla.

Kolmas vaihe sisältää konkordanssirivien tarkastelun. Se tehtiin alkuvaiheessa hakukomennoilla, joilla konkordanssitiedoista saatiin esiin rivit, joilla haluttu sane esiintyy. Myöhemmin tein analyysit uudestaan Excel-taulukkolaskentaohjelmalla, koska se tarjoaa helpon tavan tallentaa ja tulokset omille riveilleen ja työkalut rivien lajitteluun ja laskentaan.

Neljännessä vaiheessa varsinainen tutkimusaineisto ja ensimmäinen verrannaisaineisto jäsennettiin uudestaan Turun NLP-ryhmän jäsentimellä verbien sanaluokkatiedon tuottamiseksi tilastollista menetelmää varten.

Prosessin yhteenveto:

- * tekstikorpuksen kerääminen, tallentaminen ja dokumentointi
- * kevytjäsennys, lemmatisointi, virheanalyysi
- * osakorpus ja testiajot
- * tilastolliset ajot ja tutkimuskohteiden valinnat
- * tutkimuskohteiden lukumäärien kartoitus aineistoista
- * yhdyssana-analyysi ja aineistojen vertailu
- * konkordanssien ajo ja tulosten siirto taulukkolaskentaohjelmaan
- * konkordanssirivien tarkistus ja epäkelvojen hakutulosten poistaminen
- * esiintymien laskenta positioittain ja aineistojen välinen vertailu
- * määrällisten tulosten tilastollinen analyysi (osa)
- * tutkimustulosten analyysi ja tutkimuskohteiden arviointi (yhdyssanojen ja johdannaisten rooli)
- * tutkimustulosten vertaaminen verrannaisaineistoista saatuihin lähinnä määrällisiin tuloksiin
- * dependenssijäsennys, sanaluokkatiedot

5.5 Tilastolliset menetelmät

Kieliaineistojen tilastollisia menetelmiä hyödyntävässä tutkimuksessa käytetään samoja menetelmiä kuin tilastotieteessä yleensäkin (ks. esim. Rayson (2002): 39, Manning ja Schütze (1999), Oakes (1998), Evert (2005), Gries (2015), katso myös luku 2.2.1 yllä), mutta niiden soveltamisen haasteena on täsmällisesti omaan tutkimukseen soveltuvan menetelmän valinnan vaikeus. Omassa tutkimuksessani tilastollisia menetelmiä on hyödynnetty WordSmithTools-ohjelmassa, Ted Pedersenin avoimen lähdekoodin BSP/NSP-ohjelmiston tuottamia tilastollisia tunnuslukuja, Khiin neliö-testiä ja paikoitellen Fisherin eksaktia kaksipuoleista testiä. Tässä alaluvussa käsitelen näiden lisäksi Korp-konkordansserin sanakuvaa ja tilastollista analyysiä aineistojen vertaamisen tukena sekä pohdin virheiden mahdollisuutta ja niiden vaikutusta tutkimustuloksiin.

WordSmithTools

Tutkimuksen alkuvaiheen analyysit tehtiin WordSmithTools-ohjelmalla, joka tuotti tekstikorpuksesta avainsana-analyysin (keyness) ja hakusanan yhteisesiintyjät aineistosta halutun kokoisesta tutkimusikkunasta joko logaritmisen uskottavuusosamäärän (log likelihood) tai keskinäisinformaation (Mutual Information) pohjalta viiden esiintymän kynnyksarvolla. Ohjelman log likelihood -laskennan perusteella yhdistelmä *toimenpide+toteuttaa* osoittautui tilastollisesti merkitseväksi kollokaatioksi EU-asetusten aineistossa. (Westerlund (2000))

WordSmithTools-ohjelmalla tehty alkuvaiheen tilastollinen ajo ratkaisi tutkimuskohteen valinnan. Vaikka alkuvaiheen analyysijä ei ole tässä tutkimuksen lopputuloksessa hyödynnetty, analyysin tulos vaikutti tutkimuksen kulkuun siten, että kun *toimenpide* nousi myöhemmissäkin testeissä korkealle ranking-listalla, oli sen valinta yhdeksi tutkimuskohteeksi motivoitu. Myöhemmät tilastolliset ajot on tehty BSP/NSP-ohjelmistolla (ks. alaluku 5.5). Yhdistelmä *toimenpide+toteuttaa* nousi esiin myös osana BSP:n tuottamia tilastollisesti merkitsevissä yhteisesiintymissä. BSP:n käytön tavoitteena oli myös varmistaa, että alkuvaiheen työ ja erityisesti sen tuottamat johtopäätökset voidaan myöhempien tulosten valossa arvioida uudelleen.

BSP/NSP

Ted Pedersenin⁵⁰ avoimen lähdekoodin ohjelma Bigram Statistic Package v0.4 (BSP, Banerjee ja Pedersen (2003)) laskee tilastollisia tunnuslukuja bigrammin eli kahden

⁵⁰University of Minnesota Duluth, <http://www.d.umn.edu/~tpederse>

kielellisen yksikön muodostamista esiintymistä. Nämä yhteisesiintymän sisäistä assosiaatiota arvioivat menetelmät ovat Log-likelihood ratio, Pearsons chi-squared test, Fisher's exact test (Left-sided tai Right-sided), Mutual Information ja Dice Coefficient. Pedersen korvasi BSP:n vuonna 2004 päivitetyllä ohjelmalla Ngram Statistic Package⁵¹, jossa tilastollisten menetelmien valikoiman kasvamisen lisäksi kielellisten yksiköiden muodostamien esiintymien osien lukumäärä ei enää ole rajoitettu kahteen yksikköön (Pedersen (2011)).

Tein aluksi BSP:n avulla listauksen kolmen tilastollisen menetelmän tuloksiin perustuvan ranking-listan koko tutkimusaineistosta, jossa tulokset on järjestetty menetelmien Left-sided Fisher's exact test, Mutual Information ja Log-likelihood ratio laskemien tunnuslukujen perusteella optimaaliseen järjestykseen. Tilastollisesti merkitsevien yhteisesiintymien lista oli huomattavan pitkä, sillä iso määrä esiintymiä ylitti ohjelmassa oletusarvona olevan merkitsevyyden kynnyksarvon, joka tuolloin oli viisi esiintymää.

BSP:n käytön tavoitteena tutkimuksessa oli selvittää, ovatko alkuvaiheessa jo tutkittu substantiivi *toimenpide* ja sen yhteisesiintyjien joukosta tutkittavaksi valitut verbit myös BSP:n tulosten perusteella motivoitavissa. BSP tuotti merkitsevistä kahden elementin yhteisesiintymistä 60 000 rivin pituisen listauksen, joista noin 45 000 ensimmäistä riviä saivat ranking-luvun 1 ja merkitsevyyksluvun 1.0000 (Fisher's (Left-sided) exact test). Tulokset sisältävät kaikki kielelliset elementit, myös välimerkit ja lyhenteet, mikä lisää tulosten määrää huomattavasti. Vasta 45 000 tulosrivin jälkeen nelinumeroisessa desimaalissa alkaa hienoinen muutos. Myöhemmin Pedersen ehdotti kymmenen desimaalin tarkkuutta, jolloin eroavaisuuksia näkyisi jo aikaisemmin, mutta tutkimuksen kannalta riitti, että *toimenpide* ja tutkimusverbit löytyivät ranking-listan alkupäästä.

Sen jälkeen ajoin uudelleen varsinaisesta tutkimusaineistosta listauksen kolmen kielellisen elementin kokoisesta ikkunasta (noodi +/- 1 esiintymä) käyttäen pelkästään Fisherin eksaktia testiä (Fisher's (Left-sided) exact test) ja etsin tuloslistalta ne rivit, joilla sane *toimenpide* esiintyi. Kaikkien havaintojen merkitsevyyksarvo on 1.0000. Testit tehtiin varsinaisella tutkimusaineistolla, ja sen tavoitteena oli laskea aineiston perusteella bigrammien esiintymistajuuksia ja arvioida kahden sanan todennäköisyyttä esiintyä yhdessä todennäköisemmin kuin erikseen. Jälkikäteen ajateltuna olisi ollut parempi valita testiksi esimerkiksi Fisherin kaksipuoleinen eksakti testi, mutta tuolloin valinta tehtiin kuitenkin tässä esitetyn tuloslistan perusteella ja tilastollisen testin valinnassa käytettiin BSP-sivustolta löytyneitä ohjeita.

Taulukko 8 sisältää 15 ensimmäistä riviä kyseisestä listauksesta. Listauksen perusteella päätös valita *toimenpide* tarkemmin tutkittavaksi substantiiviksi vahvistui.

⁵¹<http://ngram.sourceforge.net>

Taulukko 8. Tulostiedoston 15 ensimmäistä riviä; ohjelma BSP; tilastomenetelmä LeftFisher; hakusana toimenpide; syöte EU-aineisto. Lukuohje: bigrammin I kielellinen yksikkö, <>, bigrammin II kielellinen yksikkö, <>, tunnusluvut ja merkitsevyysarvo + kielellisten yksiköiden I ja II yhteisesiintymien lukumäärä + I kielellisen yksikön esiintymien lukumäärä aineistossa + II kielellisen yksikön esiintymien lukumäärä aineistossa.

toimenpide	<>, <>	1	1.0000	337	2181	79467
toimenpide	<>olla<>	1	1.0000	245	2181	40013
säätää	<>toimenpide<>	1	1.0000	189	3990	2181
tarvittava	<>toimenpide<>	1	1.0000	172	574	2181
toteuttaa	<>toimenpide<>	1	1.0000	133	1633	2181
tarkoittaa	<>toimenpide<>	1	1.0000	127	6076	2181
toimenpide	<>ei<>	1	1.0000	108	2181	8975
tämä	<>toimenpide<>	1	1.0000	107	13217	2181
toimenpide	<>soveltaa<>	1	1.0000	83	2181	6069
toimenpide	<>toteuttaa<>	1	1.0000	69	2181	1633
toimenpide	<>se<>	1	1.0000	68	2181	19384
tarpeellinen	<>toimenpide<>	1	1.0000	59	449	2181
kyseinen	<>toimenpide<>	1	1.0000	51	3786	2181
eri	<>toimenpide<>	1	1.0000	45	471	2181
tasoitustoimenpide	<>alainen<>	1	1.0000	45	59	376

Tutkittavien verbien valintaan vaikutti WordSmithTools-ohjelman tuottama kollokaatio *toteuttaa toimenpide* varsinaisen tutkimusaineiston analyysin tuloksissa sekä verbien *toteuttaa* lähisynonyymeiltä vaikuttavien *ryhtyä, soveltaa, suorittaa, säättää* ja *tehdä* esiintyminen tutkimuskohteen *toimenpide* tekstiyhteydessä (Westerlund (2000)).

Kaikki tutkittavaksi otetut sanayhdistelmät eivät ole NSP:n tulosten perusteella tilastollisesti merkitseviä yhteisesiintymiä. Kaikki sen tuottamien potentiaalisten elementtien komponentit eivät myöskään sijaitse samassa kielellisessä yksikössä ja tulokset sisältävät paljon ylimääräisiä epäkelpoja tuloksia. Tilannetta arvioidakseni tein EU-aineistolla kuuden yhdyssanan (*erityistoimenpide, suojatoimenpide, täytännönpanotoimenpide, turvallisuustoimenpide, turvatoimenpide* ja *valvontatoimenpide*) potentiaalisten yhteisesiintymien vertailun siten, että tarkistin NSP:n tuottamista tuloksista tutkimusverbit käsin 11 saneen konkordansseista. Tämän otoksen perusteella näyttää siltä, että NSP:n tuloksissa on kaikkien tutkittujen yhdyssanojen tuloksissa huomattavasti enemmän potentiaalisia yhteisesiintymiä kuin niitä todellisuudessa löytyy. Vertailun tulokset on esitetty taulukossa 9.

Tilastollisten menetelmien rooli tutkimuskohteiden valinnassa oli siis selvittää, löytyvätkö tutkimuskohde *toimenpide* ja tutkimusverbit tilastollisesti merkitsevien yhteisesiintyjien joukosta. Esimerkiksi verbi *tehdä* ei täytä tilastollisen merkitsevyyden ehtoa esiintyessään hakusanan *toimenpide* jälkeen aineistossa, koska se esiintyy aineistossa muutoinkin kuin *toimenpiteen* yhteydessä niin taajaan (vrt. Sinclairin merkitsevät ja tavalliset kollokaatiot alaluvussa 2.2.4). Tuloslistan rivi näyttää tältä:

toimenpide tehdä 0.0536 3 4262 2181

Tutkimusverbi *tehdä* esiintyy hakusanan *toimenpide* oikealla puolella (jälkeen) aineistossa kolmessa tapauksessa, kun *tehdä* esiintyy koko aineistossa 4262 kertaa ja *toimenpide* 2181 kertaa (myöhemmin tehdyssä tarkistetussa konkordanssianalyysissä frekvenssi on 2180). Luku 0.0536 on tilastollinen merkitsevyyssarvo. Kaiken kaikkiaan verbi *tehdä* esiintyy hakusanan *toimenpide* kanssa samassa kielellisessä yksikössä 19 kertaa. Se on kuitenkin mukana tutkimusverbien joukossa, koska se on mahdollinen vaihtoehto muille tutkituille verbeille. Verbi *tehdä* esiintyy ensimmäisessä verrannaisaineistossa (SL) ja sen voisi olettaa olevan kääntäjälle tuttu potentiaalisenä yhteisesiintyjänä, koska se kokemukseni perusteella on tavallinen yleiskielessä ja esimerkiksi lääketieteen ammattikielessä: potilaalle *tehdään toimenpiteitä*. Intuitiivinen arvio lääketieteen kielestä saa tukea konkordanssityökalu Korpilla⁵² tehdystä hausta Kielipankin korpuksen *1990- ja 2000-luvun suomalaisia aikakaus- ja sano-*

⁵²corp.csc.fi

Taulukko 9. NSP-ohjelman tulosten vertailua aineistosta tuotettujen konkordanssien esiintymien välillä kuuden yhdyssanan osalta.

	EU NSP	EU tarkistettu
erityistoimenpide	soveltaa 19 soveltaminen 3 suorittaa 1 tehdä 2 toteuttaa 12 toteuttaminen 0	soveltaa /soveltaminen 12 suorittaa 0 tehdä 0 toteuttaa 6 toteuttaminen 2
suojatoimenpide	tehdä 5 toteuttaa 13 soveltaa 2 soveltaminen 5	tehdä 0 toteuttaa 11 soveltaa 0 soveltaminen 1
täytäntöönpano- toimenpide	soveltaa 1 suorittaa 1	soveltaa 0 suorittaa 0
turvallisuustoimenpide	toteuttaa 1	toteuttaa 1
turvatoimenpide	sovelluttaminen 1 soveltamisala 1 soveltaa 5	sovelluttaminen 1 soveltaminen 3 soveltaa 3
valvontatoimenpide	soveltama 2 soveltaa 9 suorittaminen 1 suorittaa 1 tehdä 3 tehtävä 1 toteuttaa 12	soveltama 0 soveltaa 0 säättää 9 suorittaa 0 tehdä 0 tehtävä 0 toteuttaa 7

*malehtiä*⁵³ osakorpuksen *Tiedelehtiä* joukosta valituista kuudesta lääketieteen alan lehdestä (*Kliinlab*, *Kognitiivinen psykoterapia*, *Psykologia-lehti*, *Solubiologi*, *Sosiaalilääketieteellinen Aikakauslehti* ja *Sydänääni*, yhteensä 3 354 572 sanetta). Niissä tutkimusverbien esiintymien jakaumat 11 sanan tutkimusikkunassa hakusanan *toimenpide* tekstiyhteydessä ovat *ryhtyä* 1, *tehdä* 58, *toteuttaa* 8, *soveltaa* 0, *suorittaa* 7 ja *säätää* 0. Verbin *tehdä* esiintymistajuuteen yleiskielessä palaan tutkimusosiossa 6.4 ja Korp on esitelty lyhyesti alla.

On huomattava, että tilastolliset testit BSP/NSP-ohjelmilla tehtiin hyvin varhaisessa vaiheessa tutkimusta, ilman perusteellista ymmärrystä tilastollisista menetelmistä. Lopputuloksen kannalta valinnat ovat kuitenkin osoittautuneet toimiviksi. Verbin *tehdä* lisäksi tutkimuskohteeksi valikoitui verbi *säätää*, joka ilman ohjelman tuottamaa listausta olisi karsiutunut verbien joukosta sen suuren esiintymistajuuden ja todellisten esiintymien suhteellisen vähyyden vuoksi.

Yli puolessa *toimenpiteen* esiintymistä sekä EU-aineistossa että SL-aineistossa on yhteisesiintyjänä jokin tutkimusverbeistä (tarkemmin luvussa 7).

Konkordanssityökalu Korpin sanakuva

Kielipankin aineistojen tarkasteluun on käytettävissä konkordanssityökalu Korp. Sen avulla on voi tehdä aineistosta rajoitettuja hakuja, laajennettuja hakuja ja edistyneitä hakuja. Rajoitetun haun puolella Korp tuottaa konkordanssirivien lisäksi dependenssisuhteisiin perustuvan sanakuvan, josta löytyvät hakusanan tavallisimmat etu- ja jälkimääritteet sekä sen edellä ja jäljessä esiintyvät tavallisimmat verbit. Esimerkkinä on hakusanan *toimenpide* sanakuva osakorpuksesta Virallinen lehti (ks. kuva 5.1). Laajennetun haun puolella onnistuu haun kohteena olevan aineiston rajoittaminen aineiston teksti- tai rakenneattribuuttien perusteella. Hakukriteeriksi voi valita esimerkiksi perusmuotoisen tai pintamuotoisen sanan, haun voi kohdistaa tiettyyn tekstilajiin tai rajata ajanjaksoa aineiston sisältämien attribuuttien sallimissa rajoissa. Korpin käyttöohje löytyy Kielipankin sivustolta⁵⁴.

Tutkimuksen toinen (semfinlex-lait), kolmas (LaDi) ja neljäs verrannaisaineisto (lehdet-osa) ovat Kielipankin aineistoja ja Korpin avulla tutkittavissa.



Tilastollinen analyysi aineistojen vertaamisen tukena

Tutkimukseni lähtökohtana oli aluksi absoluuttinen ja suhteellinen frekvenssitieto havaintokohteista ja oletus, että se tarjoaa riittävän tarkan välineen tutkimusaineis-

⁵³urn:nbn:fi:lb-2017091901

⁵⁴<https://www.kielipankki.fi/support/korp/>

Suomi | Ruotsi | Muut kielet | Rinnakkaistekstit

 Virallinen lehti valittu — 30,61M / 8,74G sanetta 

Yksinkertainen Laajennettu Edistynyt Vertailu

toimenpide (substantivi) Hae ▾

myös alkuosa loppuosa ja samaista pien- ja suuraakkoset

virkkeistä ▾ jotka sisältävät:

Konkordanssi: osumia sivulla: 25 ▾ järjestä korpuksen sisällä: järjestämätön ▾ Tilastoja: laskentaperuste: sana (+2)

Konkordanssi Tilastoja Sanakuva Kartta

Etumääritte		toimenpide	Jälkimääritte	toimenpide	verbi	Verbi	toimenpide
1. suorittaa	8224	1. antaa	105	1. tehostaa	2	1. hyväksyä	201
2. viranomainen	2752	2. ryhtyä	14	2. liittyä	5	2. käynnistää	17
3. tehdä	844	3. suunnitella	11	3. eudelle	1	3. toteuttaa	25
4. konkurssihallinto	61	4. saavuttaa	7	4. edetä	2	4. aiheuttaa	16
5. tarvita	66	5. estää	6	5. aiheuttaa	3	5. sisältää	19
6. edellyttää	39	6. todeta	5	6. konkretisoida	1	6. ehdottaa	19
7. aiheuttaa	28	7. suoritett	3	7. ävät	1	7. kohdistaa	5
8. aiheutua	75	8. pyrkiä	5	8. stavat	1	8. käsitellä	61
9. tähänastinen	16	9. päättää	5	9. mahdollistaa	2	9. listata	3
10. peruuttamaton	7	10. päästä	4	10. tähdätä	1	10. koota	5
11. tarpeellinen	21	11. realisointi	5	11. suorita	1	11. määritellä	7
12. palvelu	12	12. vähentää	4	12. istua	1	12. kontrolloida	2
13. vesihuoltosuunnitelma	6	13. varmistaa	4	13. edistää	2	13. nopeuttaa	3
14. edistää	12	14. käsitellä	5	14. tarkastella	1	14. kiirehtiä	2
15. konkreettinen	8	15. edistää	4	15. edellyttää	5	15. teettää	2

Kuva 5.1: Sanakuva hakusanana *toimenpide*, dependenssijäsennyksen perusteella konkordanssin etu- ja jälkimääritteet sekä verbikollokaatit lehdessä Virallinen lehti. Hakusanan esiintymien yhteismäärä 14806.

tosta tehtyjen havaintojen vertaamiseen verrannaisaineistoihin. Tilastollisia menetelmiä on kuitenkin käytetty täydentämään frekvenssitietoa ja tuottamaan lisätietoa jatkotutkimusta varten.

Tutkimuksessa hyödynnetään paikoitellen tilastollisia menetelmiä varsinaisesta tutkimusaineistosta (EU) ja ensimmäisestä verrannaisaineistosta (SL) tehtyjen havaintojen välisten erojen tilastollisen merkitsevyyden selvittämiseksi. Lisäksi on verrattu toisiinsa varsinaista tutkimusaineistoa ja toista verrannaisaineistoa (semfinlex-lait). Tavoitteena on tutkia 1) esiintyykö tutkimuskohde *toimenpide* tutkimusaineistossa ja verrannaisaineistossa yhtä usein ja 2) esiintyykö tietty tutkimusverbi tutkimuskohteen kanssa yhteisesiintymässä tutkimusaineistossa ja verrannaisaineistossa yhtä usein. Testataan, onko aineistojen välillä havaittu ero suurempi kuin mitä muuttujien arvojen perusteella voisi odottaa. Tämä ilmaistuna hypoteesein kuuluu seuraavasti:

Nollahypoteesi = Tutkimuskohteen *toimenpide* suhteellinen osuus esiintymistä testatun tutkimusverbin kanssa samassa yhteisesiintymässä verrannaisaineistossa on yhtä suuri kuin tutkimusaineistossa. Muuttujat ovat toisistaan riippumattomia, eikä verrattavien aineistojen välillä ole eroa tässä suhteessa.

Vaihtoehtoinen hypoteesi = Tutkimuskohteen *toimenpide* suhteellinen osuus esiintymistä testatun tutkimusverbin kanssa samassa yhteisesiintymässä verrannaisaineistossa on eri kuin tutkimusaineistossa. Muuttujat ovat toisistaan riippuvaisia ja aineistojen välillä on ero.

Aineisto on normaalisesti jakautunut, kun mediaani ja keskiarvo eivät eroa toisistaan huomattavasti. Parametrisia testejä voi käyttää, kun niiden taustaoletukset täyttyvät: muuttujat noudattavat normaalijakaumaa, ne ovat luokitteluasteikollisia tai järjestysasteikollisia ja havaintoja on riittävästi (> 30). Jos kieliaineisto olisi normaalisesti jakautunut, tilastolliseen testaamiseen voisi käyttää esimerkiksi parametristä t-testiä. Kieliaineistojen todennäköisyysjakauma on kuitenkin vino (Zipfin laki (Manning ja Schütze (1999): 24), mikä kuvastaa kielessä harvoin tai vain kerran esiintyvien elementtien paljoutta. Kielestä tehdyt havainnot eivät ole välimatka-asteikollisia, eivätkä tässä tutkimuksessa myöskään järjestysasteikollisia. Näistä syistä tämän tutkimuksen muuttujien testaamiseen paremmin soveltuvia testejä ovat parametrittomat testit, jotka eivät perustu olettamuksiin aineiston todennäköisyysjakaumasta tai muuttujien jaottelun asteikosta. Tähän tutkimukseen soveltuvia parametrittomia testejä ovat Khiin neliötesti ja Fisherin eksakti testi.

Tavoitteena on selvittää, voidaanko osoittaa sellaista tilastollista eroa esiintymistapaajuuksien välillä eri aineistoissa, jossa voidaan hyväksyä vaihtoehtoinen hypoteesi. Ellei voida, nollahypoteesi jää voimaan. Tilastollisen merkitsevyyden rajana

on viiden prosentin todennäköisyys, jolloin merkitsevyytaso tai riskitaso (level of significance) eli p-arvo on < 0.05 . Todennäköisyys havaita ero otoksessa satunaisotannalla nolla-hypoteesipopulaatiosta on $< 5\%$.

Tutkimuksen muuttujat

Ristiintaulukon/frekvenssitaulukon rivit ovat esiintymistiedon muuttujia: havaintokohde esiintyy tai ei esiinny. Tässä tutkimuksessa on käytetty nelikenttätestiä, jossa verrataan kahta aineistoa toisiinsa, joten sarakemuuttujina ovat vertailun kohteena olevat aineistot. Tutkimuksen muuttujat ovat luokitteluasteikollisia: joko verbi esiintyy tutkimuskohteen *toimenpide* kanssa samassa yhteisesiintymässä kyseisessä aineistossa tai se ei esiinny. Frekvensseistä on käytetty absoluuttisia arvoja.

Khiin neliö -testi

Khiin neliö -testi mittaa riippumattomuutta vertaamalla odotusarvoja havaintoarvoihin (suuri p-arvo tarkoittaa sitä, että havaitut erot odotusarvojen ja havaintoarvojen välillä ovat pienet ja ne voivat hyvinkin johtua sattumasta, jolloin muuttujien välinen riippuvuus ei luonnollisestikaan ole tilastollisesti merkitsevää eikä nollahypoteesiä voida hylätä). Jos saadaan tulokseksi pieni merkitsevyytaso eli p-arvo (< 0.05), erot ovat suuret ja todennäköisyys, että tulokset johtuvat sattumasta on pieni. Tällöin voidaan hylätä nollahypoteesi ja todeta, että on todennäköistä, että vaihtoehtoinen hypoteesi pitää paikkansa. Selvitetään, millä todennäköisyydellä olemme väärässä, jos päättellemme, että vaihtoehtoinen hypoteesi pitää paikkansa. Khiin neliötesti perustuu havaintojen ja havaintojen odotusarvojen eroon ja näistä laskettujen kertymäarvojen summan suhteeseen teoreettiseen Khiin neliön jakumaan. Testin toimivuus paranee sitä mukaa, kun aineiston koko kasvaa. Testin käyttö edellyttää, että havainnoista koostetun ristiintaulukon/frekvenssitaulukon jokaisen solun odotusarvo on vähintään viisi.

Tässä tutkimuksessa osa tutkimustuloksista on testattu R-ohjelman (Baayen (2008)) Khiin neliötestillä ja milloin näin on tehty, ilmoitetaan testin tulos tutkimustuloksen analyysin yhteydessä. Tutkimuksen taulukoiden solujen arvot on testattu ja kaikki odotusarvot ylittävät raja-arvon. Testimuuttuja on dikotominen: verbi esiintyy tutkimuskohteen *toimenpide* kanssa yhteisesiintymässä tietyssä aineistossa tai ei esiinny. Tutkimustuloksista selviää, miten usein tutkittava verbi esiintyy tutkimusaineistossa (EU) tai jossakin verrannaistaineistoista (SL, semfinlex-lait ja lehdet-osa) tutkimuskohteen kanssa ja miten usein verbi esiintyy kyseisessä aineistossa. Muuttuja on tieto siitä, esiintyykö verbi tutkimuskohde *toimenpiteen* kanssa yhteisesiintymässä vai ei, jolloin sen oletetaan esiintyvän aineistossa muussa ym-

Taulukko 10. Ristiintaulukko, jossa solujen arvot ovat frekvenssejä.

Havainnot	Odotusarvot			
	EU	SL	EU	SL
sane on toimenpide	2180	2763	1374	3569
sane ei ole toimenpide	1369299	3559229	1370105	3558423

päristössä. Varsinaisesta tutkimusaineistosta (EU) saatuja tuloksia verrataan ensimmäiseen verrannaisaineistoon (SL), toiseen verrannaisaineistoon (semfinlex-lait) ja neljanteen verrannaisaineistoon (lehdet-osa). Kolmas verrannaisaineiston (LaDi) poikkeaa sisällöllisesti muista aineistoista, joten sitä koskevat havainnot on käsitelty erillisinä aineiston luonne huomioiden.

Taulukossa 10 esitän esimerkin ristiintaulukosta, joka sisältää saneen *toimenpide* esiintymät tutkimusaineistossa ja ensimmäisessä verrannaisaineistossa. Niiden perusteella on laskettu Khiin neliö -testin avulla merkitsevyysarvo.

Tutkitaan, millä todennäköisyydellä voidaan hyväksyä vaihtoehtoinen hypoteesi, jonka mukaan *toimenpide* esiintyy molemmissa aineistoissa yhtä todennäköisesti. Esimerkissä Khiin neliö -testin ($X\text{-squared} = 655.24$, $df = 1$, $p\text{-value} < 2.2e-16$) perusteella voidaan todeta, että riskitaso, jolla tehdään virhe, jos hylätään nollahypoteesi on minimaalinen, jolloin vaihtoehtoista hypoteesia voidaan pitää erittäin todennäköisenä.

Virheiden mahdollisuus

Miten varautua virheisiin tilastollisessa testauksessa? Virhe voi johtua siitä, että hylätään nollahypoteesi, vaikka se pitäisi hyväksyä. Silloin syynä voi olla, että tulos johtuukin satunnaisvaihtelusta (sattumasta) eikä merkitsevästä erosta, tai on käytetty väärää testiä. Toisessa virhetyypissä hyväksytään nollahypoteesi, vaikka se pitäisi hylätä. Tällainen virhe voi johtua aineiston pienuudesta tai laadusta – tai on käytetty väärää testiä. Tutkimuksessa pyritty varmistamaan, että virhettä ei tapahdu selvittämällä testin lähtöoletukset aineiston kannalta ja testi on valittu parametrittomien testien joukosta. Tiedetään myös, että valittu testi on yksisuuntainen, koska verrataan verrannaisaineiston muuttujia tutkimusaineiston muuttujiin eikä tutkita muuttujien yhteisvaikutusta.

Tilastolliset testit on tehty avoimessa⁵⁵ R-ohjelmointiympäristössä⁵⁶. Käytetty testi ja sen tulos on ilmoitettu luvussa 6 testatun muuttujan määrällisten havaintojen yhteydessä.

Empiirisen osan rajanvetoa

Ajattelen, että yhteisesiintyminen ilmiönä on jatkumo, enkä edellytä sen esiintymistäajuuden ylittävän ennalta määrättyä kynnyсарvoa. Näin jatkumon molemmat päät ovat teoreettisesti osa ilmiön kenttää. Käytännössä ilmiön toteutumaa saadaan näkyville isoista aineistoista vain kieliteknologisin menetelmin, jotka perustuvat elementtien laskemiseen. Avainsana-analyyseissä oli tutkimuksen alkuvaiheessa oletusarvona kynnyсарvo, joka jo siinä kohtaa vaikutti tutkimuskohteen valintaan.

Tulosten analysoinnissa olen pyrkinyt huomioimaan myös yksittäiset esiintymät siitä huolimatta, että yksittäisellä esiintymällä ei katsota olevan todistusvoimaa. Säädoskielen teksteissä ei kuitenkaan ole samanlaista lähtökohtaista vapautta tai pyrkimystä välttää toistoa kuin yleiskielessä tai mahdollisesti joissakin muissa erikoiskielissä tai ammattikielissä olisi. Vaikka tutkimuksen lähtökohta ja pääasiallinen tavoite on kartoittaa aineistojen välisiä eroja määrällisten havaintojen kautta, jotka toisinaan edellyttävät tilastollisen merkitsevyyden testaamista, ovat yksittäiset valtavirrasta poikkeavat havainnot osoitus kielen variaatiosta ja heterogeenisuudesta. Erikoiskielisen aineiston osalta jää tutkimuksen johtopäätöksissä harkittavaksi, onko tilanne tutkimuskohteena olevan erikoiskielen kannalta toivottava. Tarkastelen tutkimuskohteena olevien yhteisesiintymien toistuvuutta varsinaisessa tutkimusaineistossa ja sen ensimmäisessä verrannaisaineistossa lähtökohtaisesti absoluuttisten ja suhteellisten esiintymistäajuuksien avulla ja teen niiden pohjalta tutkimusaineiston ja verrannaisaineistojen välisen vertailun.

Yhteisesiintymisen puhtaan tilastollisen näkökulman merkitystä tutkimuksessa on vaikea arvioida. Distributionaalinen tarkastelu edellyttää tilastollista vahvistusta siitä, että tulokset ovat riippumattomuudesta sattumasta. Toisaalta taas yksittäiset tapahtumat voivat tuottaa tietoa kääntäjän tekemästä tietoisesta tai tiedostamatomasta valinnasta. Aineistossa harvinainen tapahtuma voi siten tarjota sellaista tietoa käännösprosessista ja kääntämisen kohteesta, joka muutoin jäisi pimentoon.

⁵⁵<https://www.gnu.org/licenses/gpl-3.0.html>

⁵⁶<https://www.r-project.org/>

5.6 Tekstikorpuksen käsittelyn työkalut

Tutkimus on lähtökohtaisesti määrällinen. Sen ensisijaisena tavoitteena on tarkastella tutkimuskohteiden esiintymistä varsinaisessa tutkimusaineistossa (EU) ja verrata tilannetta ensin ensimmäiseen verrannaisaineistoon (SL). Saatuja tuloksia verrataan toiseen verrannaisaineistoon (semfinlex-lait), kolmanteen verrannaisaineistoon (LaDi) aineiston luonne huomioiden ja neljäänteen verrannaisaineistoon (lehdet-osa). Tutkimuskohteiden kanssa samassa kielellisessä yksikössä esiintyvien tutkimusverbiin määrät ja sijainnit positioissa tutkimusikkunassa selvitetään kieliteknologisten työkalujen avulla. Tässä alaluvussa esittelen laskennassa käytetyt ohjelmat sekä aineistojen käsittelyn työkalut, konkordansserit, ohjelmat ja jäsentimet.

Aineistojen saneiden tai sanaesiintymien (*token*) ja esiintymätyyppien (*type*) määrät varsinaisesta tutkimusaineistosta (EU) ja ensimmäisestä verrannaisaineistosta (SL) laskettiin unix-komennoilla. Kun *toimenpide* oli valittu tutkimuskohteeksi ja ensimmäiset ajatukset yhdyssanoista kollokaatioiden ilmentyminä heräsivät, verrattiin aineistoja toisiinsa Miikka Silfverbergin kirjoittaman skriptin *vertaa.pl* avulla. Tavoitteena oli selvittää, mikä aines on aineistoissa uniikkia ja mikä yhteistä. Näin pystyttiin esimerkiksi tuottamaan listat hakusanan *toimenpide* sisältävistä yhdyssanoista ja selvittämään, mitkä yhdyssanoista esiintyvät molemmissa aineistoissa ja mitkä vain toisessa. Tietoa hyödynnettiin myöhemmin yhdyssanojen tarkastelussa itsenäisenä kollokaationa tai yhteisesiintymän osana.

Konkordanssityökalut

Tutkin hakusanan käyttöympäristöä tuottamalla perusmuotoistetusta aineistosta konkordansseja, joihin ohjelma tulostaa tutkittavan elementin lisäksi viisi sanetta sen molemmilta puolilta (Stubbs (1996): 175–176). Tällaisen kwic-analyysin (*key word in context*) tai konkordanssianalyysin tekemiseen löytyy nykyisin useita ilmaisia sovelluksia. Tätä tutkimusta varten on tehty kaksi konkordanssiria. Perusmuotoistetusta aineistosta tuotetut konkordanssit on tehty Miikka Silfverbergin kirjoittamalla Perl-ohjelmilla. Pintamuotoisesta aineistosta tuotetut konkordanssit on tehty Jonas Westerlundin Javalla kirjoittamalla konkordanssityökalulla. Molemmat ajetaan komentoriviltä ja niiden tulostiedostot ohjataan tekstitiedostoiksi, joista voidaan halutessa tehdä hakuja tai laskea rivien (ts. saneiden) lukumääriä tavanomaisilla unix-työkaluilla, tai tuloksen voi viedä taulukkolaskentaohjelmaan käsiteltäväksi tarkemmin. Tutkimusikkunan koon voi molemmissa ohjelmissa valita, mutta koon kasvattaminen tuottaa enemmän käsin analysoitavaa materiaalia. Analyysiä varten tuotetut konkordanssit sisältävät tutkittavan hakusanan lisäksi positioissa L5–L1 ja R1–R5 sijaitsevat kielelliset elementit, viisi sanetta ennen ja jälkeen hakusanan.

Tutkimuksen toinen verrannaisaineisto (semfinlex-lait), kolmas verrannaisaineisto (LaDi) sekä neljäs verrannaisaineisto (lehdet-osa) ovat käytettävissä Kielipankin kautta. Niiden haut on tehty Ruotsin kielipankin Språkbankenin kehittämällä konkordanssityökalulla **Korp**⁵⁷, joka on FIN-CLARIN-hankkeessa⁵⁸ muokattu Suomen Kielipankin tarpeisiin. Löydettyään hakusanan Korp näyttää kyseisen hakusanan sisältävän virkkeen. Menetelmä eroaa varsinaisen tutkimusaineiston (EU) ja ensimmäisen verrannaisaineiston (SL) menetelmistä seuraavasti:

1. Hakukomennossa tai sen yhteydessä ei määritellä erikseen, miten monta kielellistä yksikköä hakusanan molemmin puolin ohjelma sisällyttää konkordanssiin, koska tuloksena on aina kokonainen virke. Varsinaisen tutkimusaineiston (EU) ja ensimmäisen verrannaisaineiston (SL) hauissa konkordanssityökaluilla tutkimuskunan kooksi on määritelty +/- 5 kielellistä yksikköä. Korp-työkalun tulostiedostosta on mahdollista tarkastella laajempaa tekstiyhteyttä. Erikoiskielen aineistossa ominaisuus on hyödyllisempi kuin mitä tutkimuksen alkuvaiheessa osasin arvella.

2. Hakukomennon antamisen yhteydessä määritellään, haetaanko aineistosta tulostettava virke hakusanan perusmuodon vai sananmuodon perustella.

3. Tulostiedostoon virkkeen sisältämät kielelliset elementit tulostuvat omille riveilleen, jolloin laadullisen tutkimuksen puolella tehtävä arviointi helpottuu, mutta toisaalta laskenta ja yhteisesiintyjien lajittelu positioihin on hankalampaa verrattuna omilla konkordanssereilla tuotettuun tulokseen.

4. Kolmannen verrannaisaineiston (LaDI) eri tekstilajit (direktiivit, asetukset ja lait) sisältyvät samaan aineistoon, koska aineiston attribuuteissa ei ole mukana tietoa tekstilajista. Koska aineisto on huomattavasti varsinaista tutkimusaineistoa ja ensimmäistä verrannaisaineistoa pienempi, on lajittelu tekstilajeihin ollut mahdollista tehdä käsin Excelissä. Tulokset ovat absoluuttisia frekvenssejä, koska tietoa aineiston sisältämien tekstilajien koosta ei ole saatavissa.

Korp laskee rajoitetun haun avulla tuotetun konkordanssin perusteella aineistosta sanakuvan, joka sisältää hakusanan kanssa tavallisimmin esiintyvät etu- ja jälkimääritteet sekä verbit. Sanakuvasta pääsee hakutuloksesta käsin tarkastelemaan konkordanssirivejä ja lähdetiedostoja. LaDi-aineistossa eri tekstilajit eivät kuitenkaan ole eriteltävissä, jolloin sanakuvan tulokset sisältävät esiintymiä kaikista tekstilajeista. Sen sijaan neljännen verrannaisaineiston (1900- ja 2000-luvun lehdet) tarkastelussa sanakuva osoittautui hyväksi vaihtoehdoksi käsitellä tuloksia tulosten suuren määrän vuoksi. Tutkituista verbeistä taajimmin kyseisessä aineistossa esiintyy verbi *tehdä*, jonka edustajia aineistossa on 350 509 kappaletta. *Tehtä*-verbin sanakuva oli tulosten suuren määrän vuoksi tuotettava useammassa jaksossa, jolloin

⁵⁷ www.kielipankki.fi

⁵⁸ www.finclarin.fi

näille yhteistä sanakuvaa ei ole tuotettavissa. Sanakuvan tuloksia onkin tutkimuksessa hyödynnetty vahvistamaan tutkimustuloksia tai tuottamaan niille vastakkaista todistusaineistoa.

Jäsenin Turku Neural Parser Pipeline (TNPP)

Turku Neural Parser Pipeline⁵⁹ on Turun yliopiston NLP-ryhmän kehittämä neuroverkkojäsenin, joka on avoimen lähdekoodin ohjelmana vapaasti saatavissa. Käytän jäsentimestä Kielipankin tavoin lyhennettä TNPP. TNPP on Universal Dependencies -hankkeen työhön perustuva yli 50 kieltä jäsentävä dependenssijäsenin, joka on ollut ladattavissa NLP-ryhmän sivustolta keväästä 2019 alkaen. Siinä vaiheessa tutkimusta oli selvää, että tulosten arviointi tilastollisin menetelmin on paikoin tarpeen ja vuonna 1999 perusmuotoistettu tutkimusaineisto (EU) ja sen ensimmäinen verrannaisaineisto (SL) jäsennettiin TNPP-jäsentimellä. Tavoitteena ei enää siinä vaiheessa tutkimusta ollut dependenssirelaatioiden hyödyntäminen analyysissä, johon erityisesti aineiston erityispiirteistä, virkkeiden pituudesta ja tutkimuskohteen ominaispiirteistä sekä jo tehdystä manuaalisesta analyysistä. Tutkimustulokset olivat jo valmiina, mutta tilastollista analyysia varten tarvittiin kuitenkin tieto eri verbien lukumäärien lisäksi myös siitä, miten monta verbiä aineistot sisältävät. Onkin huomioitava, että tilastollinen analyysi perustuu kahdesta lähteestä peräisin oleviin lähtötietoihin ja eri jäsentimillä käsitellyt aineistot poikkeavat aina jonkin verran toisistaan.

Taulukkolaskentaohjelma tulosten lajittelussa

Kaikki konkordanssien tuottamat tulokset siirrettiin taulukkolaskentaohjelmaan käsiteltäviksi. Tulokset järjestettiin positioittain (sarakkeittain) aakkosjärjestykseen, jolloin tutkittavat yhteisesiintyjät on suhteellisen helppo laskea ja taulukoida. Jokainen konkordanssi lajiteltiin uudelleen kymmenen kertaa jokainen sarake erikseen, jotta positioissa toistuvat rakenteet saatiin lasketuksi. Sen jälkeen esiintymät laskettiin ja tulokset taulukoitiin exceliin. Toistuvat lausekkeet käsiteltiin erikseen, milloin se osoittautui tarkoituksenmukaiseksi – esimerkiksi säännöllisesti toistuvassa vakiolauseessa voi olla yhteisesiintyjien lisäksi muuttuvia komponentteja, vaikkapa hakusanaa määrittävä adjektiivi.

⁵⁹<https://turkunlp.org/Turku-neural-parser-pipeline/>

Yhteenveto

Tutkimuksessani käytetyt ohjelmat käyttötarkoituksineen on koottu alla olevaan taulukkoon 11. Taulukosta selviää, mihin tarkoituksiin ohjelmaa on käytetty. Yleisesti ottaen menetelmiä on käytetty käsittelemään aineistoa, tuottamaan frekvenssitietoa, tuottamaan aineistosta konkordansseja ja laskemaan tilastollisia tunnuslukuja.

Taulukko 11. Tutkimuksessa käytetyt ohjelmat.

Menetelmä	Tavoite
WordSmithTools	1998 avainsana-analyysi, konkordanssit, toimenpi*-sanalistat, kollokaatit
UNIX-komento: wc uniq	rivien lukumäärä = kielellisten yksiköiden määrä; sanamäärät; merkkimäärät; aineistojen koot, sanaesiintymät (token)/esiintymätyypit (type)
UNIX-komento: grep	yksittäisen saneen esiintymät tai esimerkiksi tietty pintamuoto tai lisärajoitteilla tai säännöllisillä lausekkeilla mikä tahansa tulosjoukko
konkordanssi.pl	tuottaa konkordanssirivit hakusanan (lemma) perusteella perusmuotoistetusta aineistosta (perl-ohjelma)
konkordanssi.java	tuottaa konkordanssirivit hakusanan (lemma) perusteella pintamuotoisesta aineistosta (java-ohjelma)
vertaa.pl	verrata kahden tiedoston sisältöjä keskenään, tulostaa rivit, jotka eivät ole yhteisiä molemmille tiedostoille
Bigram Statistics Package, Ngram Statistics Package	selvittää tilastollisten menetelmien avulla, onko tietyn hakusanan valinnalle tutkimuskohteeksi perusteita sekä mitä mahdollisia kollokaatteja voisi hakea tekstiyhteydestä
Korp	tuottaa konkordanssit toisesta, kolmannesta ja neljänestä verrannaisaineistosta
TNPP	tuottaa tutkimusaineiston ja ensimmäisen verrannaisaineiston jäsenyyksen tilastollista analyysia varten
R	tuottaa tutkimusaineiston ja ensimmäisen verrannaisaineiston tulosten tilastollisen analyysin

5.7 Yhdyssanojen kartoituksen menetelmä

Yhdyssanoja on etsitty tutkimusaineistosta (EU) ja ensimmäisestä verrannaisaineistosta (SL) kahdella tavalla. Miikka Silfverbergin kirjoittama skripti tuotti aineistoista rivit, joissa haluttu sane tai lemma esiintyy osana yhdyssanaa, sen lopussa, alussa tai keskellä. Näin saatiin listat lemmatusta aineistosta ja tekstimuotoisesta aineistosta sellaisista yhdyssanoista, jotka sisältävät halutun hakusanan. Listat ajettiin kahdeksasta hakusanasta, joiden joukossa olivat myös tutkimuskohde *toimenpide*.

Yhdyssanojen konkordanssit on tuotettu Jonas Westerlundin kirjoittamalla konkordansserilla. Yhdyssanat esiintymisfrekvensseineen taulukoitiin Excelliin sen mukaan, missä kohtaa hakusana sijaitsee yhdyssanassa. Tämän tutkimuksen kannalta tärkeimmässä yhdyssanatyypissä hakusana on yhdyssanan edusosa, kuten esimerkiksi 2 alla, jossa esitetään osa hakusanan *toimenpide* tulostiedostosta hakusanan *toimenpide* sijaitessa yhdyssanan lopussa (SL).

- (2) aluekehitystoimenpiteiden
edistämistoimenpiteet
edistämistoimenpiteisiin
edistämistoimenpiteisiin
edistämistoimenpiteitä

5.8 Tutkimusikkunan laajuus

Kollokaatiotutkimuksen alkuvaiheissa kollokaatiotutkimus käsitti käytännössä hakusanan lähiympäristön, joka useimmiten tarkoitti hakusanaa ja sen molemmin puolin esiintyvää yhtä ortografista sanaa (Sinclair (1991)). Jantunen totesi tutkimuksessaan adjektiiveista *tärkeä* ja *keskeinen* (Jantunen (2001)), että kaksi sanaa hakusanan molemmin puolin käsittävä tutkimusikkuna (tutkimusalue) ei näiden adjektiivien osalta juurikaan tarjoa määrällisesti enempää tuloksia kuin vieruskollokaatit (yksi sana hakusanan molemmin puolin) käsittävä ikkuna. Synonymiaa ja käännösuomea sekä syntaksin että leksikon kannalta käsittelevässä väitöstutkimuksessaan Jantunen (2004) käyttää tarkastelualueena ikkunaa, jonka koko on kolme esiintymää ennen noodia (hakusanaa) ja kaksi esiintymää sen jälkeen. Tutkimusikkunan kokoon vaikuttaa paitsi tekstilaji, myös tutkimuskohde.

Englanninkielistä aineistoa tutkiva Sinclair päätyi suosittelemaan ikkunaa, jossa noodin ympäriltä tutkitaan molemmin puolin neljän kielellisen elementin verran kontekstia, vaikkakin hänen havaintojensa mukaan suurin osa havainnoista sijoittuu

heti noodin molemmin puolin (Sinclair et al. (2004)). Päätökseen ikkunan koosta vaikuttavat tutkittavan aineiston tekstilaji ja käytettävissä olevat analyysimenetelmät. Säädoskieli on siinä mielessä vaativa tutkimuskohde, että virkkeet ja lauseet ovat usein pitkiä ja lausekkeissa. Kun käytännössä yhdistetään sekä vaatimus isosta ikkunasta että halutaan sallia myös kielessä tavallisempien saneiden tutkimus sen sijaan, että keskityttäisiin erikoiskielen harvinaisempaan ainekseen, tulevat ajot raskaiksi ja tuloslistat pitkiksi ja manuaalisen työn osuus kasvaa helposti mahdottomaksi.

Arppe esitti suomen kieltä käsittelevässä tutkimuksessaan vuodelta 2008, että useimmat kollokaatit (dependentit) esiintyvät ikkunassa, jonka laajuus on +/- 5 sanetta hakusanan ympärillä Arppe (2008).

Tutkittavat positiot on merkitty seuraavasti:

Hakusana

L5–L1 Hakusanaa ennen tekstissä esiintyvät yksiköt; hakusanan vasemmalla puolella olevat viisi yksikköä kymmenen yksikön tutkimusikkunassa

R1–R5 Hakusanan jälkeen tekstissä esiintyvät yksiköt; hakusanan oikealla puolella olevat viisi yksikköä kymmenen yksikön tutkimusikkunassa

Omassa tutkimuksessani olen valinnut tutkimusikkunan kooksi viisi esiintymää ennen ja jälkeen hakusanan. Tutkimusikkunan koko on siten 11 sanetta. Varsinaisessa tutkimusaineistossa konkordanssi voi sisältää rivejä, joissa positioissa L5–L1 ja R1–R5 on välimerkkejä, puolipisteitä, pilkkuja tai pisteitä. Tarkastelen myös, osuuko tutkimusikkunan sisään tapauksia, joissa noodi ja kollokaatti (hakusana ja yhteiseesiintyjä) sijaitsevat eri virkkeissä ja voivatko ne muodostaa yhteiseesiintymän siitä huolimatta, etteivät sijaitse samassa lausekkeessa.

5.9 Epäkelvot hakutulokset

Konkordanssien rivien lukumäärät ja aitojen esiintymien lukumäärät on laskettu manuaalisesti taulukkolaskennalla. 'Aito esiintymä' tarkoittaa tässä konkordanssia, jossa jokin tutkimusverbeistä esiintyy hakusanan kanssa samassa kielellisessä yksikössä, tavallisesti lausekkeessa. Näiden erotuksesta selviää, miten paljon rivejä on siirretty poistettujen rivien tiedostoon eli miten paljon on perusmuotojen perusteella ollut mahdollista määritellä epäkelvoiksi hakutuloksiksi. Jos haku on tehty perusmuotoisesta aineistosta, on joitakin konkordanssirivejä tavallisesti tarkistettava pintamuotoisesta aineistosta, koska aina ei ole mahdollista tulkita millaisessa suhteessa hakusana ja sen laskennallinen yhteiseesiintyjä ovat toisiinsa. Seuraavassa on tällaisesta tilanteesta kaksitulkinainen esimerkki:

- (3) ryhtyä muu tämä laki säättää **toimenpide**. <p> Elintarvikevirasto voida määrätä

Esimerkin pintamuoto voisi olla joko 'ryhtyä muuhun tässä laissa säädettyyn toimenpiteeseen' tai 'ryhtyä muuhun tämän lain säätämään toimenpiteeseen' ja näiden monikkomuodot (SL). Jälkimmäinen olisi tämän tutkimuksen kriteerein kelvollinen yhteisesiintymä.

5.10 Yhteenvedo ja tutkimusosion rakenne

Olen edellä luvussa 2 hahmotellut kollokaation jakautuvaksi kahdeksi eri ominaispiirteiden avulla määritettäväksi ilmiöksi. Yhtäältä se on distributionaalinen, määrällinen tapahtuma kielessä, jonka esiintymiä on mahdollista laskea tekstistä tai puheesta ja arvioida näiden esiintymien vakiintumisen astetta kyseisessä kontekstissa. Tätä tapahtumaa kutsun yhteisesiintymäksi. Vaikka esiintymistäajuus kuvastaa yhteisesiintymän tavallisuutta, on myös yksittäinen havainto yhteisesiintymän instantaatio, joka sattumoisin kyseisessä aineistossa esiintyy vain kerran mutta jolla on potentiaalia jossakin toisessa aineistossa olla taajemmin esiintyvä. Yhteisesiintyjä on siten kontekstisidonnainen käsite. Toisaalta kollokaatio voidaan nähdä merkitykseen sitoutuvana ilmiönä, joka voi ilmetä vakiintuneena mutta myös harvinaisena tapahtumana. Tätä tapahtumaa kutsun vakiintumaksi. Tarkastelemalla distribuutiota pyrin löytämään aineistosta tietoa vakiintumasta ja keinoja hahmottaa sen ominaispiirteitä konkreettisemmin. Näin kollokaatio toimisi paremmin kääntäjän työkaluna ja sen potentiaali sekä yhteisesiintymisen että vakiintumisen puolelta tulee näkyvämmäksi ja paremmin hyödynnetyksi.

Seuraavissa luvuissa 6 ja 7 esitän tutkimuskohdetta *toimenpide* koskevat havainnot. Analyysiosio pyrkii peilaamaan yhteisesiintymän ominaispiirteitä, jotka on esitelty tarkemmin luvussa 2. Ne ovat määrälliset ominaispiirteet (toistuvuus ja tyyppisyys), rakenteeseen liittyvät ominaispiirteet (osien välinen etäisyys/läheisyys, sisäkkäisyys, lauseke ja vakiolauseke), merkitykseen liittyvät piirteet (semanttinen preferenssi ja semanttinen prosodia, kompositionaalisuus) ja assosiaatioon liittyvät piirteet (vastavuoroisuus, konventionaalisuus ja norminmukaisuus). Havaintojen lähtökohtana on laskenta: esiintyykö tutkimuskohteen *toimenpide* tekstiympäristössä jokin mielenkiinnon kohteena olevista verbeistä *ryhtyä*, *soveltaa*, *suorittaa*, *säättää*, *tehdä* tai *toteuttaa*.

Luku 6 kattaa määrälliset ja rakenteeseen liittyvät havainnot tutkimuskohteesta (alla olevan listan kohdat 1–12) ja luku 7 merkitykseen ja assosiaatioon liitty-

vät havainnot (kohdat 13–16). Tutkimuksen tavoitteena on verrata yhteisesiintymiä varsinaisessa tutkimusaineistossa (EU) ja ensimmäisessä verrannaisaineistossa (SL). Tuloksia verrataan kolmeen muuhun aineistoon (semfinlex-lait, LaDi ja lehdet-osa). Toinen verrannaisaineisto tarjoaa mahdollisuuden tarkastella yhteisesiintymiä ajallisessa perspektiivissä ja laajemman säädösaineiston. Kolmas verrannaisaineisto tarjoaa lisätietoa tutkimuskohteena olevien EU:n asetusten rinnalla Suomen lainsäädäntöön integroitavien direktiivien osalta ja neljäs verrannaisaineisto yleiskielisestä aineistosta.

Määrälliset havainnot jakautuvat seuraavasti:

1. tutkimuskohde tutkimusaineistossa
2. tutkimusverbit yhteisesiintyjinä
3. positiot ja yhdyssanat on raportoitu myös rakenteen yhteydessä

Rakenteeseen liittyvät havainnot jakautuvat seuraavasti:

4. komponentit: yhdyssanat erikoistapauksena
5. tekninen rakenne: havainnot kollokaateista positioissa L5-L1 ja R1-R5
6. tekninen rakenne: havainnot hakuikkunan ulkopuolelta
7. tekninen rakenne: yhteisesiintymän elementtien lukumäärä
8. tekninen rakenne: useita verbejä samassa rakenteessa erikoistapauksena
9. tekninen rakenne: vakiolausekkeet erikoistapauksena
10. kieliopillinen rakenne: verbikantaiset substantiivit
11. kieliopillinen rakenne: epäsuorat objektit hakutuloksissa
12. kieliopillinen rakenne: kieliopillinen luku

Merkitykseen ja assosiaatioon liittyvät havainnot jakautuvat seuraavasti:

13. tutkimuskohteen vastineet IATE-termitietokannassa
14. merkityksen muutos tai tarkentuminen
15. semanttinen preferenssi ja semanttinen prosodia
16. tekstiyhteys ja tilannekonteksti

Luku 6

Määrälliset ja rakenteeseen liittyvät havainnot: *toimenpide*

Tässä luvussa esitän tutkimuskohteen *toimenpide* ja sen kanssa hakuikkunassa samassa syntaktisessa rakenteessa esiintyvien tutkimusverbien *ryhtyä, soveltaa, suorittaa, säätää, tehdä ja toteuttaa* lukumäärät ja rakenteeseen liittyvät havainnot. Havainnot on kerätty varsinaisesta, EU-asetukset sisältävästä tutkimusaineistosta (EU), Suomen säädöskokoelman teksteistä kootusta ensimmäisestä verrannaisaineistosta (SL), toisesta verrannaisaineistosta, joka on osa Semanttinen Finlex-aineistoa (semfinlex-lait), kolmannesta verrannaisaineistosta Lakeja ja direktiivejä, Kielipankki-versio (LaDi) sekä neljännessä verrannaisaineistosta, joka on osa Kielipankin 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä -korpusta (lehdet-osa). Aineistot on kuvattu tarkemmin luvussa 4.

Tutkimusaineisto (EU) sisältää Euroopan unionin asetuksia ja sitä verrataan ensimmäiseen verrannaisaineistoon (SL), joka edustaa suomalaisen säädöskielen normia varsinaisen tutkimusaineiston koostamisen ajalta, 1900 ja 2000 -lukujen vaihteesta. SL-aineisto edustaa sen aikaista odotuksenmukaista tapaa ilmaista asioita. Tuloksia on verrattu kahteen lakikieltä sisältävään korpukseen, joista Kielipankin Semfinlex-aineisto sisältää Finlex-lakikokoelman alkuperäiset lait (aineisto on julkaistu Kielipankissa vuonna 2019) ja korpus Lakeja ja direktiivejä, Kielipankki-versio (LaDi) koostuu laeista, direktiiveistä ja neuvoston asetuksista sekä neuvoston ja parlamentin asetuksista. Asetukset integroidaan suoraan voimassa olevaan lainsäädäntöön, kun taas direktiivit hyväksytään tavallisen lainsäädäntömenettelyn kautta, jossa niitä muokataan osana prosessia. Neljäs verrannaisaineisto (lehdet-osa) taas tarjoaa mahdollisuuden tarkastella tutkimuskohdetta yleiskielisessä aineistossa. Tutkimuskysymykset ja käsitteet on esitelty tarkemmin luvussa 1, tutkimuskohteiden valinta aineistoluvussa 4 ja menetelmä luvussa 5.

Etsin vastauksia ensimmäiseen tutkimuskysymykseen, joka on koko työn ydin. Kysyn, onko EU-asetusten suomentamisen seurauksena syntynyt epätyypillisiä tai uusia yhteisesiintymiä. Ensisijainen vastaus tähän kysymykseen tässä tutkimuksessa on määrällinen. Painopiste on siten yhteisesiintymien lukumäärissä ja suhteellisissa esiintymistajuuksissa, joiden perusteella pyrin arvioimaan, miten samanlaisia tai erilaisia tutkimusaineisto ja ensimmäinen verrannaisaineisto tässä suhteessa ovat. Vertaan saatuja havaintoja toiseen, kolmanteen ja neljänteen verrannaisaineistoon. Tutkimuskohteesta *toimenpide* esitetään tässä luvussa määrälliset analyysit sen yhteisesiintymisestä kuuden verbin kanssa, jotka ovat *ryhtyä*, *soveltaa*, *suorittaa*, *säättää*, *tehdä* ja *toteuttaa*. Tarkastelen tässä luvussa myös näiden yhteisesiintymien rakenteeseen liittyviä kysymyksiä. Tavoitteena on myös selvittää tutkimuskohteen *toimenpide* yhteisesiintymistä tutkimusverbien kanssa tässä tutkimuksessa esiteltyjen ominaispiirteiden tarkentamiseksi. Erityistä huomiota on kiinnitetty yhdyssanoihin ja niiden tarjoamiin haasteisiin yhteisesiintymän määrittämisessä.

Tavoitteena on ensisijaisesti selvittää, miten usein tutkimusverbit esiintyvät hakusanan *toimenpide* lähiympäristössä (+/- 5 ortografista yksikköä hakusanan molemmin puolin) laskemalla verbien esiintymät pintamuotoisista konkordanssiriveistä positioissa L5–L1 ja R1–R5 silloin, kun verbi ja hakusana sijaitsevat V + N -rakenteessa, jonka funktio on toiminnan ja sen kohteen ilmaiseminen.

Laskennassa varsinaisessa tutkimusaineistossa EU ja ensimmäisessä verrannaisaineistossa SL on otettu huomioon myös verbikantaiset substantiivit silloin, kun ne toimivat yhteisesiintyjänä rakenteessa (esim. *soveltaminen*) sekä yhdyssanat, joissa verbikantainen substantiivi esiintyy määräiteosana (esim. *soveltamisajanjakso*). Mainittakoon tässä yhteydessä kuitenkin, että olen sisällyttänyt verbikantaiset substantiivit tuloksissa verbin esiintymiin. Aineistojen EU, SL, semfinlex-lait ja LaDi laskenta ja epäkeltojen rivien poistaminen on tehty käsin, kun taas lehdet-aineiston analyysit on tehty hyödyntäen lisäksi Kielipankin konkordanssityökalu Korpin sanakuvaa.

6.1 Määrälliset havainnot (EU ja SL)

Tässä alaluvussa käsittelen tutkimuksen keskeiset tutkimustulokset. Niiden avulla pyrin hahmottamaan, sisältävätkö kääntämisen tuloksena syntyneet EU-asetusten suomenkieliset versiot yhteisesiintymiä, joissa tutkimuskohteen *toimenpide* kanssa on valikoitunut esiintyväksi jokin tutkimusverbeistä eri tavoin kuin Suomen säädös-kokoelman sisältämien lakien tarkastelun perusteella voisi odottaa. Tulokset ovat määrällisiä ja ne on paikoitellen validoitu tilastollisen testauksen avulla (testi ja hypoteesit on esitelty kappaleessa 5.5). Havaintoja tarkastellaan kollokaation ominaispiirteiden valossa, joiden kuvaus löytyy luvusta 2. Rakenteeseen liittyvät havainnot

Taulukko 12. Hakusanan *toimenpide* tunnusluvut määrällisten tulosten vertailussa.

	EU	SL
1) koko aineiston saneiden lukumäärä	1371479	3561992
1b) normalisointikerroin	0,073	0,028
2) kaikki saneen <i>toimenpide</i> esiintymät aineistossa	2180	2763
2b) normalisoituna per 100 000 esiintymää; esiintymistäajuus	159	77
3) kaikkien tutkimuskohteena olevien kuuden verbin kaikki yhteisesiintymät hakusanan <i>toimenpide</i> kanssa	1142	1450
3b) normalisoituna per 100 000 esiintymää	83	41

limittyvät osin määrällisiin havaintoihin ja esitän ne määrällisten havaintojen yhteydessä aineistoittain. Merkitykseen ja assosiaatioon liittyviä havaintoja käsittelemme luvussa 7.

Havaintojen tilastollista merkitsevyyttä on tarkasteltu erikseen. Aineistoja verrataan keskenään ja arvioidaan, esiintyykö tutkimusverbi tutkimuskohteen kanssa samassa yhteisesiintymässä kyseisessä aineistossa tai sen tekstiyhteydessä. Nollahypoteesi on seuraava: Tutkimuskohteen *toimenpide* esiintyminen tutkimusverbin kanssa samassa yhteisesiintymässä tutkimusaineistossa ja verrannaisaineistossa on sattumanvaraista. Menetelmä on kuvattu kohdassa 5.5.

6.1.1 Tutkimuskohde *toimenpide*

Tarkastelen tässä alaluvussa saneen *toimenpide* lukumääriä tutkimusaineistossa (EU) ja verrannaisaineistossa (SL). Tutkimuskohteesta ja tutkimusverbeistä eri aineistoista tehtyjä havaintoja verrataan toisiinsa normalisoituina suhteessa 100 000 esiintymään taulukossa 12 esitetyn normalisointikertoimen avulla.

Toimenpide on huomattavasti tavallisempi EU-aineistossa kuin SL-aineistossa (taulukko 13). EU-aineistossa saneita *toimenpide* on 2 180 kappaletta noin 1,1 miljoonan sanan korpuksessa. SL-aineistossa saneen *toimenpide* esiintymiä on 2 763 noin 3,5 miljoonan sanan korpuksessa. Karkeasti arvioiden EU-aineistossa on noin kaksinkertainen määrä lekseihin *toimenpide* esiintymiä verrattuna SL-aineistoon. Khiin neliötestin tulos ($X^2 = 655.24$, $df = 1$, $p\text{-value} < 2.2 \cdot 10^{-16}$) vahvistaa,

että nollahypoteesi, jonka mukaan *toimenpide* esiintyy yhtä usein molemmissa aineistoissa, voidaan selkeästi hylätä ja frekvenssien ero aineistoissa on tilastollisesti merkitsevä.

Aineistot eroavat toisistaan myös tutkittavaksi valittujen verbien *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä*, ja *toteuttaa*, toisin sanoen yhteisesiintyjäehdokkaiden, suhteen. Yli 50 prosentissa saneen *toimenpide* esiintymistä EU-aineistossa ja SL-aineistossa sen tekstiyhteydessä havaitaan jokin kuudesta tutkimusverbistä. Tutkimuskohteen *toimenpide* selvästi tavallisin yhteisesiintyjä tutkittavien verbien joukossa EU-aineistossa on *toteuttaa*, jonka osuus on lähes 60 prosenttia. SL-aineistossa tavallisin tutkituista yhteisesiintyjistä yli 65 prosentin osuudella tapauksista on *ryhtyä* ja seuraavaksi yleisin on verbisuurittaa, joita on yli 20 prosenttia yhteisesiintymistä. EU-aineistossa vastaavasti toiseksi yleisin verbi yhteisesiintymässä on *soveltaa* ja kolmanneksi yleisin *säätää*. Tutkittavien kuuden verbin joukossa verbin *tehdä* esiintymiä on kummassakin aineistossa vain vähän.

Tarkemmat luvut on esitetty taulukossa 13. Monet SL-aineiston havaintojen joukossa olevat EU-aineistolle tyypilliset esiintymät osoittautuvat EU-säädöksistä peräisin oleviksi tai ne ovat Euroopan neuvoston säädöksiin liittyvistä teksteistä.

Myös *toimenpiteen* käänkösvastineiden määrästä IATE-termitietopankissa voi päätellä, että se on EU-kontekstissa erittäin produktiivinen. Mahdolliset käänkösvastineet on lueteltu taulukossa 52.

Tutkimukseen valittuja verbejä voi oletettavasti löytyä myös tutkimusikkunan ulkopuolelta, erityisesti säädöskielessä, jossa virkkeet ovat tyypillisesti pitkiä. Tutkimusikkunaksi valitun alueen (5 sanetta hakusanan molemmilla puolilla) tarkempi tarkastelu esiintymien lukumääristä eri positioissa osoittaa kuitenkin, että verbiyhteisesiintyjien määrä vähenee etäännyttäessä hakusanasta, mikä näyttäisi vahvistavan ajatusta siitä, että suurin osa havainnoista löytyy hakusanan *toimenpide* lähiympäristöstä. EU-asetuksissa ja Suomen säädöskokoelman teksteissä on kuitenkin eroavaisuuksia siinä, missä positioissa tarkasteltavia verbejä esiintyy eniten. EU-asetuksissa eniten havaintoja verbistä *soveltaa* on positioista L2–L1 ja R1–R3 ja niistä suurin osa on positiossa R1. Verbin *säätää* lähes kaikki havainnot ovat positiossa L1. SL-aineistossa kumpaakaan verbiä ei juuri esiinny, ja aineiston ehdottomasti tavallisin verbi *ryhtyä* on positiossa L4 yli kaksi kertaa tavallisempi kuin positiossa L1, josta on seuraavaksi eniten havaintoja. Verbi *ryhtyä* esiintyy myös uloimpana hakusanasta/noodista olevissa positioissa, ja sen esiintyminen tutkimusikkunan ulkopuolella onkin kaikista tutkimuksen verbeistä todennäköisintä (positiokohtaiset lukumäärät löytyvät taulukoista 14 ja 15). Positiokohtainen analyysi 11 saneen tutkimusikkunassa ei kuitenkaan tue ajatusta siitä, että kauimpana syntaktisen suhteen pääsanasta lähtökielessä sijaitseva kollokaatti tuottaisi herkemmin epätyypillisen yhteisesiintymän.

Taulukko 13. Yhteenvedo hakusanan *toimenpide* kanssa tutkimusikkunassa +/-5 sanetta/yksikköä noodin ympärillä tarkasteltavien verbien esiintymistä ja niiden osuudet hakusanan *toimenpide* kaikista esiintymistä. (EU ja SL).

toimenpide EU <i>toimenpide +</i>	2 180 EU kpl	per 100 000 normali- soitu	haku- sanoista	%	% verbi- kollokaateista
<i>ryhtyä</i>	40	2,9		1,8	3,5
<i>soveltaa</i>	198	14,4		9,1	17,3
<i>suorittaa</i>	23	1,7		1,0	2,0
<i>säätää</i>	206	15,0		9,4	18,0
<i>tehdä</i>	19	1,4		0,9	1,6
<i>toteuttaa</i>	656	47,8		30,1	57,4
	1142	83,2		52,3	100,00
toimenpide SL <i>toimenpide +</i>	2 763 SL kpl	per 100 000 normali- soitu	haku- sanoista	%	% verbi- kollokaateista
<i>ryhtyä</i>	970	27,2		35,1	66,9
<i>soveltaa</i>	5	0,1		0,2	0,3
<i>suorittaa</i>	300	8,4		10,8	20,7
<i>säätää</i>	17	0,5		0,6	1,2
<i>tehdä</i>	68	1,9		2,4	4,7
<i>toteuttaa</i>	90	2,5		3,2	6,2
	1450	40,6		52,3	100,00

6.1.2 Tutkimusverbit yhteisesiintyjinä

Esitän seuraavassa havainnot tutkimusverbeistä yhteisesiintyjinä. Verbin perusmuodon lukumäärä aineistossa ei aina vastaa tutkimustulosten perusteena olevien esiintymien lukumäärää, koska yhteisesiintyjäksi on laskettu myös verbikantaisest substantiivit, joita ei tavata kaikkien verbien tuloksissa. Raportoin tutkimusverbit aakosjärjestyksessä: *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä*, *toteuttaa*. Esitän myös verbien lukumäärät positioittain, vaikka positiot sisältyvät teoriaosioissa yhteisesiintyjän rakenteen tarkasteluun.

Taulukko 14. Tilastotiedot verbien yhteisesiintymistä hakusanan *toimenpide* kanssa: niiden kappalemäärät eri positioissa, prosenttiosuudet suhteessa saneen *toimenpide* kaikkien esiintymien lukumäärään EU-aineistossa ja prosenttiosuudet tutkittavien verbien ja hakusanan *toimenpide* kaikista yhteisesiintymistä.

	L5	L4	L3	L2	L1	toimenpide	R1	R2	R3	R4	R5
ryhtyä	2	0	7	11	11	toimenpide	5	1	0	3	0
soveltaa	4	3	1	23	7	toimenpide	111	19	14	12	4
suorittaa	2	1	2	6	6	toimenpide	4	2	0	0	0
säätää	1	6	0	8	189	toimenpide	0	0	0	0	2
tehdä	0	6	0	6	1	toimenpide	0	0	0	0	6
toteuttaa	36	33	73	204	173	toimenpide	86	22	9	12	8
	L5	L4	L3	L2	L1	toimenpide	R1	R2	R3	R4	R5
ryhtyä	0,0	0,0	0,3	0,5	0,5	toimenpide	0,2	0,1	0,0	0,1	0,0
soveltaa	0,2	0,1	0,1	1,1	0,3	toimenpide	5,0	0,9	0,6	0,5	0,2
suorittaa	0,1	0,1	0,1	0,3	0,3	toimenpide	0,2	0,1	0,0	0,0	0,0
säätää	0,1	0,3	0,0	0,4	8,7	toimenpide	0,0	0,0	0,0	0,0	0,1
tehdä	0,0	0,3	0,0	0,3	0,1	toimenpide	0,0	0,0	0,0	0,0	0,2
toteuttaa	1,6	1,5	3,3	9,4	8,0	toimenpide	3,9	1,0	0,4	0,5	0,4
	L5	L4	L3	L2	L1	toimenpide	R1	R2	R3	R4	R5
ryhtyä	0,9	0,0	0,6	1,0	1,0	toimenpide	0,4	0,1	0,0	0,3	0,0
soveltaa	0,4	0,3	0,1	2,0	0,6	toimenpide	9,7	1,7	1,2	1,1	0,4
suorittaa	0,2	0,1	0,2	0,5	0,5	toimenpide	0,4	0,2	0,0	0,0	0,0
säätää	0,1	0,5	0,0	0,7	16,6	toimenpide	0,0	0,0	0,0	0,0	0,2
tehdä	0,0	0,5	0,0	0,5	0,1	toimenpide	0,0	0,0	0,0	0,0	0,5
toteuttaa	3,2	2,9	6,4	17,9	15,2	toimenpide	7,5	1,9	0,8	1,1	0,7

a) *ryhtyä* + *toimenpide*

EU-aineistossa V+N-rakenne *ryhtyä* + *toimenpide* esiintyy 40 kertaa. Normalisoituna rakenne esiintyy koko aineistossa 2,9 kertaa 100 000 sanaa kohden. SL-aineistosta rakenteita löytyy 970 kappaletta, toisin sanoen normalisoituna 27 yhteisesiintymää 100 000 sanaa kohden. Verbi *ryhtyä* esiintyy yhdessä saneen *toimenpide* kanssa tutkimusikkunassa SL-aineistossa lähes kymmenkertaisesti EU-aineistoon verrattuna. Yhteisesiintymän *ryhtyä* + *toimenpide* osuus kaikista tutkituista V+N-rakenteista EU-aineistossa on 3,5 prosenttia. Aineistojen ei ($X^2 = 35.872$, $df = 1$, $p\text{-value} = 2.107e-09$, joten voidaan todeta hyvin pieni p -arvo (< 0.05) -> Verbin *ryhtyä* suhteelliset osuudet tutkimuskohteen *toimenpide* tutkitussa tekstiyhteydessä poikkeavat aineistoissa toisistaan, ero on tilastollisesti merkitsevä ja riippuvuusoletus voidaan hylätä).

Taulukko 15. Tilastotiedot verbien yhteisesiintymistä hakusanan *toimenpide* kanssa: niiden kappalemäärät eri positioissa, prosenttiosuudet suhteessa saneen *toimenpide* kaikkien esiintymien lukumäärään SL-aineistossa ja prosenttiosuudet tutkittavien verbien ja hakusanan *toimenpide* kaikista yhteisesiintymistä.

	L5	L4	L3	L2	L1	toimenpide	R1	R2	R3	R4	R5
ryhtyä	42	430	37	138	178	toimenpide	52	49	18	10	16
soveltaa	0	2	0	2	1	toimenpide	0	0	0	0	0
suorittaa	8	9	10	28	64	toimenpide	116	41	16	4	4
säätää	0	1	1	0	8	toimenpide	1	0	3	2	1
tehdä	2	1	9	19	2	toimenpide	28	6	0	1	0
toteuttaa	2	0	4	24	15	toimenpide	32	5	4	4	0
	L5	L4	L3	L2	L1	toimenpide	R1	R2	R3	R4	R5
ryhtyä	1,5	15,6	1,3	5,0	6,4	toimenpide	1,9	1,8	0,7	0,4	0,6
soveltaa	0,0	0,07	0,0	0,1	0,0	toimenpide	0,0	0,0	0,0	0,0	0,0
suorittaa	0,3	0,3	0,4	1,	2,3	toimenpide	4,2	1,5	0,6	0,1	0,1
säätää	0,0	0,0	0,0	0,0	0,3	toimenpide	0,0	0,0	0,1	0,1	0,0
tehdä	0,1	0,0	0,3	0,7	0,1	toimenpide	1,0	0,2	0,0	0,0	0,0
toteuttaa	0,1	0,0	0,1	0,9	0,5	toimenpide	1,2	0,2	0,1	0,1	0,0
	L5	L4	L3	L2	L1	toimenpide	R1	R2	R3	R4	R5
ryhtyä	2,9	29,7	2,6	9,5	12,3	toimenpide	3,6	3,4	1,2	0,7	1,1
soveltaa	0,0	0,1	0,0	0,1	0,07	toimenpide	0,0	0,0	0,0	0,0	0,0
suorittaa	0,6	0,6	0,7	1,9	4,4	toimenpide	8,0	2,8	1,1	0,3	0,3
säätää	0,0	0,1	0,07	0,0	0,6	toimenpide	0,1	0,0	0,2	0,1	0,1
tehdä	0,1	0,1	0,6	1,3	0,1	toimenpide	1,9	0,4	0,0	0,1	0,0
toteuttaa	0,1	0,0	0,3	1,7	1,0	toimenpide	2,2	0,3	0,3	0,3	0,0

Jakauma positioittain: *ryhtyä* + *toimenpide*

Esitän taulukossa 16 yhteisesiintyjän *ryhtyä* esiintymien prosentuaaliset jakaumat positioittain EU-aineistossa ja SL-aineistossa.

Verbiyhteisesiintyjä *ryhtyä* on EU-aineistossa eniten positioissa L1 ja L2, joissa havaintoja on 11 kappaletta molemmissa (esimerkeissä 4 ja 6 esitän kyseiset yhteisesiintymät, joista on useampi kuin yksi havainto), mikä vastaa 27,5 prosenttia yhteisesiintymistä *ryhtyä* + *toimenpide*. Esimerkissä 7 on tyypillinen esiintymä positiossa L3.

Yhteisesiintymien verbien kasautuminen tiettyihin positioihin viittaa vakiorakenteisiin. Vaikka niitä löytyy, ovat lukumäärät pieniä. Vakiorakenteesta NEG + *ryhtyä* + *toimenpide* aineistossa on 4 havaintoa (tyypillinen esiintymä löytyy esimerkiksi

Taulukko 16. Verbin *ryhtyä* esiintymien prosenttiosuudet positioittain yhteisesiintymissä *ryhtyä + toimenpide*. (EU ja SL)

L5	L4	L3	L2	L1	toimenpide, EU	R1	R2	R3	R4	R5
5,0	0,0	17,5	27,5	27,5	toimenpide	12,5	2,5	0,0	7,5	0,0
L5	L4	L3	L2	L1	toimenpide, SL	R1	R2	R3	R4	R5
4,3	44,3	3,8	14,2	18,4	toimenpide	5,4	5,1	1,7	1,1	1,7

4). Kolme sellaista yhteisesiintymää, joissa verbi *ryhtyä* on positiossa L1, edustavat esimerkin 5 mukaisia rakenteita *saattaa + helpottaa + komissio + ryhtyä + toimenpide*. Verbin ollessa positiossa L2 suurin osa esiintymistä on esimerkin 6 rakenteen mukaisia yhteisesiintymiä *ryhtyä + A + toimenpide*, joissa adjektiivin vaihtelua rakenteessa on seuraava: *oleellinen* (1), *tarpeellinen* (1), *tarvittava* (1), *muu* (3). EU-aineistosta löytyy yksi SL-aineistossa tavallinen rakenne, jossa verbi *ryhtyä* on positiossa L5 (esimerkki 8).

- (4) ei ole riittävää aihetta ryhtyä toimenpiteisiin, vastineen antamiseen oikeutettujen henkilöiden (EU)
- (5) valitukset saattavat helpottaa komission ryhtymistä toimenpiteisiin, kun ETY:n perustamissopimuksen 85 (EU)
- (6) vuoksi oltava toimivalta ryhtyä tarvittaviin toimenpiteisiin, jos sopimuksen tai yhdenmukaistetun (EU)
- (7) Sopimusvaltioiden on ryhdyttävä kaikkiin tarpeellisiin toimenpiteisiin sen varmistamiseksi, että ne (EU)
- (8) ryhtyä mainittujen artikloiden soveltamisen edellyttämiin toimenpiteisiin ja tätä tarkoitusta varten (EU)

Verrannaisaineistossa (SL) eniten on yhteisesiintymiä *ryhtyä + toimenpide*, joissa verbi on positiossa L4. Näitä on reilu 44 prosenttia havainnoista. Suuri lukumäärä selittyy vakiolausekkeilla, joista SL-aineistossa taajimmin esiintyvä rakenne on *voida/saada + ryhtyä + laki/asetus/se + täytäntöönpano + edellyttää + toimenpide*, esimerkiksi *voidaan ryhtyä lain/asetuksen/sen täytäntöönpanon edellyttämiin toimenpiteisiin*. Kyseisen vakiorakenteen eri muotoja aineistossa on yhteensä 386 kappaletta eli lähes 90 prosentissa sellaisista yhdistelmän *ryhtyä + toimenpide*

esiintymistä, joissa *ryhtyä* on positiossa L4 - määrä on lähes 14 prosenttia kaikista hakusanan *toimenpide* esiintymistä koko SL-aineistossa. Kun *ryhtyä*-verbi esiintyy tutkimusikkunassa hakusanan *toimenpide* kanssa peräti 35 prosentissa kaikista hakusanan *toimenpide* havainnoista, onkin luonnollista, että paljon löytyy myös yhteisesiintymiä, joissa verbi on muissa ennen hakusanaa sijaitsevilla positioilla L1, L2, L3 ja L5. Verbien lukumäärät näissä positioissa ovat 178, 138, 37 ja 42. Näissäkin ryhmissä edellä kuvattu rakenne toteutuu usein erilaisina modifikaatioina, joista on esimerkkejä alla tutkimusverbin *ryhtyä* esiintyessä eri positioissa: esimerkissä 9 positiossa L1 (EU-aineisto), esimerkissä 10 positioissa L2 (EU-aineisto), esimerkissä 11 positiossa 4 (SL-aineisto, 430 havaintoa) ja esimerkissä 12 positiossa L5 (SL-aineisto).

- (9) tietojen perusteella ole aihetta ryhtyä toimenpiteisiin perustamissopimuksen 85 artiklan 1 kohdan (EU)
- (10) muita oikeuksia tai ryhtyy muihin toimenpiteisiin estääkseen jakelijoita tai käyttäjiä hankkimasta (EU)
- (11) Voidaan ryhtyä asetuksen täytäntöönpanon edellyttämiin toimenpiteisiin.
<p> El 107 L || Rivi 45692 (SL)
- (12) ryhtyä sen täytäntöönpanon kannalta tarpeellisiin toimenpiteisiin.
<p> Tällä lailla kumotaan || Rivi 3527516 (SL)

Ryhtyä on SL-aineiston ehdottomasti tavallisin hakusanan *toimenpide* yhteisesiintyjä, kun taas EU-aineistossa se on hyvin epätavallinen. SL-aineistossa se esiintyy yli 35 prosentissa *toimenpiteen* hakutuloksista. Suurimpana ison lukumäärän syynä tai selityksenä on koko hakutulostuloksen läpi kulkeva vakiofraasi *olla + (A/ADVLI) + ryhtyä + (A/ADVLI) + toimenpide + (OBJ) + N* (verbikantainen), esimerkiksi *pankin on kiireellisesti ryhdyttävä toimenpiteisiin sidotun oman pääoman kartuttamiseksi*. Vakiolausekkeiden adjektiiveissa on siis vaihtelua, esimerkkeinä mainittakoon yllä olevissa esimerkeissä esiintyvien lisäksi ainakin *asianmukainen, oleellinen, olennainen, tarkoituksenmukainen* ja *välttämätön* ja adverbistä esimerkiksi *välittömästi, viipymättä* tai *tarvittaessa*. Vastaavissa fraaseissa EU-aineistossa verbin *ryhtyä* sijaan esiintyy useimmiten verbi *toteuttaa* (ks. tarkemmin havainnot verbin *toteuttaa* esiintymistä alla).

b) *soveltaa + toimenpide*

EU-aineistossa yhteisesiintymä *soveltaa + toimenpide* esiintyy yhteensä 198 kertaa, normalisoituna se on noin 16 kappaletta per 100 000 sanetta ja noin 9 prosenttia kaikkien kuuden tutkittavan verbin ja *toimenpiteen* yhteisesiintymistä. SL-aineistossa esiintymiä on yhteensä vain 5, toisin sanoen 0,14 esiintymää per 100 000 sanaa ja vain reilu 0,3 prosenttia kaikkien kuuden tutkittavan verbin ja *toimenpiteen* yhteisesiintymistä. *Soveltaa* on siten 103 kertaa tavallisempi verbi EU-aineistossa kuin SL-aineistossa sellaisessa V + N -rakenteessa, joissa substantiivina on *toimenpide* ja 51 kertaa tavallisempi esiintymissä, joissa *toimenpide* esiintyy yhdessä jonkin tutkimusverbin kanssa. (X-squared = 227.45, df = 1, p-value < 2.2e-16, joten voidaan todeta hyvin pieni p-arvo (< 0.05) -> Verbin *soveltaa* suhteelliset osuudet tutkimuskohteen *toimenpide* tutkitussa tekstiyhteydessä poikkeavat aineistoissa toisistaan, ero on tilastollisesti merkitsevää ja riippuvuusoletus voidaan hylätä).

EU-aineistoissa yhteisesiintyjä *soveltaa* esiintyy pääosin positiossa R1, johon siijoittuu yli 55 prosenttia yhteisesiintymän *soveltaa + toimenpide* havainnoista. Seuraavaksi eniten, reilussa 11 prosentissa havainnoista, sen löytää positiossa L2 jossa se on tavallisin verbikantaisena substantiivina *soveltaminen*. Tarkemmat prosenttiluvut löytyvät taulukosta 17. SL-aineistossa havaintoja on vain positioissa L1, L2 ja L4, yksi tai kaksi havaintoa kussakin.

Taulukko 17. Verbin *soveltaa* esiintymien prosenttiosuudet positioittain yhteisesiintymissä *soveltaa + toimenpide*. (EU ja SL)

L5	L4	L3	L2	L1	toimenpide, EU	R1	R2	R3	R4	R5
2,0	1,5	0,5	11,5	4,5	toimenpide	55,5	9,5	7,0	6,0	2,0
L5	L4	L3	L2	L1	toimenpide, SL	R1	R2	R3	R4	R5
0,0	40,0	0,0	40,0	20,0	toimenpide	0,0	0,0	0,0	0,0	0,0

Yhteisesiintyjä *soveltaa* esiintyy EU-aineistossa useimmiten positiossa R1 (111 esiintymää), josta seuraavat esimerkit:

Esimerkit 13 ja 14 kuvaavat yhteisesiintyjän *soveltaa* esiintymistä positiossa R1 (EU-aineisto). Jälkimmäisessä verbikantainen substantiivi on rakenteessa *lykätä + soveltaminen*. Samanlaisia rakenteita EU-aineistossa on 19 kappaletta.

- (13) 1 kohdan c alakohdassa tarkoitettua toimenpidettä sovelletaan ainoastaan mainitun asetuksen (EU)

Psitiossa L1 verbi *soveltaa* esiintyy EU-aineistossa yllä mainitussa rakenteessa *soveltaa toimenpidettä/toimenpiteitä* seitsemässä tapauksessa yhdeksästä, joista

- (14) Siinä tapauksessa komissio lykkää päättämiensä toimenpiteiden soveltamista enintään kuukaudella ilmoituksen tekopäivästä (EU)

kahdessa tapauksessa *toimenpiteisiin sovelletaan* tiettyä lain, säädöksen tai ohjeen kohtaa tai edellytyksiä ja rajoja. Nämä havainnot eivät sisälly määrällisiin tuloksiin.

Esimerkissä 15 yhteisesiintyjä *soveltaa* on positiossa L2 (EU-aineisto).

- (15) jotka on kutsuttu soveltamaan ehkäiseviä toimenpiteitä, ja b) aikaansaadakseen (EU)

Verbi *soveltaa* on SL-aineistossa hakusanan *toimenpide* lähiympäristössä erittäin harvinainen. Se on komponenttina 15:ssä tutkimusverbin ja hakusanan *toimenpide* yhdistelmässä. Myös SL-aineistossa verbi *soveltaa* esiintyy tutkimusikkunan sisällä, mutta silloin *toimenpide* on verbin epäsuora objekti: lakeja tai määräyksiä sovelletaan *toimenpiteeseen* (tutkimusikkunassa vastaavia havaintoja oli 12 kappaletta). Nämä havainnot eivät sisälly määrällisiin tuloksiin.

Esimerkki 16 on SL-aineistosta ja verbi *soveltaa* on siinä positiossa L2. Samanlaisia esiintymiä on kaksi kappaletta.

- (16) Kansakuntien jäsenille kehoituksen soveltaa näitä toimenpiteitä.
<p> Niihin saattaa sisältyä (SL)

c) *suorittaa + toimenpide*

EU-aineistossa yhteisesiintymä *suorittaa + toimenpide* esiintyy vain 23 kertaa, toisin sanoen normalisoituna noin 2 kertaa 100 000 sanetta kohden. SL-aineistossa esiintymiä on yhteensä 300, toisin sanoen normalisoituna noin 8 esiintymää per 100 000 sanetta. SL-aineistossa on siten noin neljä kertaa enemmän kyseisen rakenteen edustajia kuin EU-aineistossa. (X-squared = 23.536, df = 1, p-value = 1.226e-06, joten voidaan todeta hyvin pieni p-arvo (< 0.05) -> *suorittaa* ei esiinny yhtä todennäköisesti *toimenpiteen* kanssa molemmissa aineistoissa, voidaan hylätä riippuvuusoletus)

EU-aineistossa verbi *suorittaa* esiintyy *toimenpiteen* lähiympäristössä tasaisesti positioissa L5 – R2, joista positioista L1 ja L2 on molemmista 6 havaintoa. Kyseisissä positioissa on siten hiukan yli neljäsosa havainnoista, toisin sanoen reilu 26 prosenttia yhteisesiintymän *suorittaa + toimenpide* esiintymistä. SL-aineistossa ylivoimaisesti eniten eli 102 kappaletta (hiukan yli 36 prosenttia havainnoista) löytyy positioista R1 ja seuraavaksi eniten esiintymiä on positiossa L1, 64 kpl eli reilu 22 prosenttia. Prosentuaalinen jakauma positioittain on esitetty taulukossa 18.

Taulukko 18. Verbin *suorittaa* esiintymien prosenttiosuudet positioittain yhteisesiintymissä *suorittaa + toimenpide*. (EU ja SL)

L5	L4	L3	L2	L1	toimenpide, EU	R1	R2	R3	R4	R5
8,7	4,4	8,7	26,1	26,1	toimenpide	17,4	8,6	0,0	0,0	0,0
L5	L4	L3	L2	L1	toimenpide, SL	R1	R2	R3	R4	R5
2,7	3,0	3,3	9,3	21,3	toimenpide	38,7	13,7	5,4	1,3	1,3

Esimerkissä 17 yhteisesiintyjä *suorittaa* on EU-aineistossa positiossa L5. *Ryhtymisen* ohella Suomen säädösteksteissä *toimenpiteitä* on usein myös *suoritettu*. Alla olevassa esimerkissä 17 EU:n asetuksessa voidaan *toimenpide* sekä *suorittaa* että *toteuttaa*. Esimerkissä 17 yhteisesiintyjä on positiossa L2 EU-aineistossa ja esimerkissä 19 vastaavasti SL-aineistossa positiossa R2. Havainto verbin *suorittaa* esiintymästä positiossa R3 löytyy esimerkistä 20.

- (17) Suorittaa tämän asetuksen mukaisesti toteutetut toimenpiteet, sekä viranomaiset ja toimielimet (EU)
- (18) tai että niille suoritetaan muita toimenpiteitä kuin niiden kunnan säilyttämiseksi tarvittavat (EU)
- (19) ,kustannusten jakaminen tai muu toimenpide on suoritettava vasta hakemuksessa tarkoitetun (SL)
- (20) Patenttiviranomaisen tulee todeta , että toimenpide on katsottava suoritetuksi määräajassa, (SL)

SL-aineistossa on yksi verbikantainen substantiivi rakenteessa *lykätä + sovelta-**minen*, kun verbiyhteisesiintyjänä on *suorittaa*, esimerkki 21 alla:

- (21) .<p> Milloin metsäkeskuksen määräämien toimenpiteiden suorittamista ei voida lykätä, (SL)

d) *säätää + toimenpide*

EU-aineistossa yhteisesiintymiä *säätää + toimenpide* on yhteensä 206, toisin sanoen havaintoja on 17 per 100 000 sanetta. SL-aineistossa esiintymiä on yhteensä 19, toisin sanoen 0,5 per 100 000 sanetta. Verbin *säätää* suhteellinen esiintymistajuus

tutkittavan yhteisesiintymän komponenttina on siis häviävän pieni SL-aineistossa verrattuna EU-aineistoon. (X-squared = 806.88, df = 1, p-value < 2.2e-16, joten voidaan todeta hyvin pieni p-arvo (< 0.05) -> Verbin *säätää* suhteelliset osuudet tutkimuskohteen *toimenpide* tutkitussa tekstiyhteydessä poikkeavat aineistoissa toisistaan, ja ero on tilastollisesti merkitsevä ja riippuvuusoletus voidaan hylätä).

Molemmissa aineistoissa verbi *säätää* on selkeästi useimmin esiintyvä yhteisesiintyjä positiossa L1, jossa EU-aineistossa on 189 kappaletta (yli 93 prosenttia esiintymistä) ja SL-aineistossa yli 8 kappaletta eli 47 prosentissa kaikista havainnoista (mutta esiintymiä SL-aineistossa positiossa L1 on siis vain 8 esiintymää). Seuraavaksi eniten verbin *säätää* esiintymiä on EU-aineistossa positioissa L2 (mutta vain 5 kappaletta eli vähemmän kuin 3 prosenttia havainnoista) ja L4 (mutta vain 6 kpl eli vajaa 3 prosenttia havainnoista). SL-aineistossa positiossa R3 on 3 kappaletta eli vähän alle 18 prosenttia havainnoista. Luvut on esitetty taulukossa 19.

Taulukko 19. Verbin *säätää* esiintymien prosentiosuudet positioittain yhteisesiintymissä *säätää* + *toimenpide*. (EU ja SL)

L5	L4	L3	L2	L1	toimenpide, EU	R1	R2	R3	R4	R5
0,5	2,9	0,0	3,9	91,8	toimenpide	0,0	0,0	0,0	0,0	0,9
L5	L4	L3	L2	L1	toimenpide, SL	R1	R2	R3	R4	R5
0,0	5,9	5,9	0,0	47,1	toimenpide	5,9	0,0	17,6	11,7	5,9

EU-aineiston positiossa L1 olevista yhteensä 189 esiintymästä 148 edustaa sellaisen vakiorakenteen modifikaatioita, joissa toimenpide **säädetään asetuksessa**. Esimerkkejä modifikaatioista löytyy taulukosta 20.

Kun verrannaisaineistossa (SL) **säädetään toimenpiteitä**, ovat kyseiset lainkohdat jäljitettävissä eurooppalaiseen lainsäädäntöön, kuten Euroopan neuvostoon liittyviin säädöksiin. EU-kontekstissa esiintymät sisältävät useimmiten yllä mainitun vakiorakenteen.

Vakiorakenteesta INE + *säätää* + *toimenpide* yli 78 prosenttia EU-aineiston esiintymistä verbin ollessa positiossa L1 on muotoa *asetus* + *säätää* + *toimenpide*. SL-aineistossa positiossa L1 on yhteensä 5 verbin *säätää* esiintymää ja näistä 3 vakiorakenteessa.

Tavallisin sisällöllinen täydennys vakiolausekkeeseen on, että säädetty toimenpide on jonkin EU:n komitean lausunnon mukainen. Komitea voi olla esimerkin 22 mukaisesti sokerin hallintokomitea, tai kyseessä voi olla jokin muu maatalouteen liittyvä hallintokomitea, esimerkiksi kalastustuotteiden, viinin, sianlihan, naudanlihan,

Taulukko 20. Vakiorakenteen modifikaatioita (inessiivi) positiossa L1, jossa kaikki verbin *säätää* esiintymät sijaitsevat. (EU ja SL)

EU		SL	
asetuksessa	148		
artiklassa	21	§:ssä	3
kohdassa	6	kohdassa	1
osastossa	2		
luetelmakohdassa	2	alakohdassa	1
säännöksissä	1	laissa	2
päätöksessä	1		
Yhteensä	181		7
pilkku, joidenkin, jokaisen edellä, edellyttävät, 1600/92	8	asetuksella	1
Kaikki L1 säätää	189		8

viljan, siementen, vihannesten tai yleisesti ottaen kyseinen tai asianomainen hallintokomitea. *Lausunnonmukaisuus* mainitaan tekstiyhteydessä 68 konkordanssirivissä.

(22) , ja tässä asetuksessa säädetyt toimenpiteet ovat sokerin hallintokomitean lausunnon mukaiset (EU)

Säätää on *toimenpiteen* yhteisesiintyjänä SL-aineistossa erittäin harvinainen (17 esiintymää). EU-asetuksissa sen suhteellinen yhteisesiintyminen *toimenpiteen* kanssa on SL-aineistoon verrattuna yli 15-kertainen. Toisaalta vaihtelu EU-asetuksissa on lähes olematonta ja lähes kaikki esiintymät verbin ollessa positiossa L1 ovat osa vakiorakennetta.

Asetuksissa **säädetyt toimenpiteet sovelletaan** (9 esiintymää) ja **toteutetaan** (10 esiintymää). Sen sijaan muita tutkimuksen kohteena olevia verbejä *tehdä* ja *suorittaa* ei ole mukana verbin vaihtelussa tutkimusikkunassa.

e) *tehdä* + *toimenpide*

EU-aineistossa *tehdä* + *toimenpide* yhteisesiintyvät yhteensä 19 kertaa, jolloin esiintymiä noin 1,5 kappaletta 100 000 sanaa kohden. Laskentatulos on verbin *tehdä* osalta erikoinen, sillä tuloksissa on konkordanssirivejä vain 7 kappaletta. Koska kaikki

esiintymät lasketaan erikseen, näyttäytyy esiintymien luku 19 suhteettoman suurena haun tulosjoukon kokoon nähden. Konkordanssiriveistä 6 ovat identtisiä ja kaikissa kuudessa verbi *tehdä* esiintyy kolme kertaa samassa konkordanssissa (ks. esimerkki 23 alla; laskentamenetelmä on esitelty tarkemmin luvussa 5). SL-aineistossa esiintymiä on 68 kappaletta, toisin sanoen normalisoituna 2 kappaletta 100 000 sanaa kohden. *Tehdä* on sekä tutkimusaineistossa (EU) että ensimmäisessä verrannaisaineistossa (SL) siis suunnilleen yhtä harvinainen verbi *toimenpiteen* yhteisesiintyjänä. (X-squared = 3.7501, df = 1, p-value = 0.05281, joten p-arvo > 0.05, jolloin riippuvuusoletus jätettävä voimaan verbin *tehdä* esiintymistä *toimenpiteen* kanssa)

Tutkittavat verbikollokaatit ovat EU-aineistossa pääosin sijoittuneet neljään positioon: L4, L2, ja R5, joissa kaikista verbiyhteisesiintyjä *tehdä* löytyy 6 kappaletta. Ne ovat kaikki kuitenkin identtisissä virkkeissä. Lisäksi positiossa L1 on yksi esiintymä. Prosenttiosuudet on esitetty taulukossa 21 ja niistä on esimerkit 23 ja 24 alla. Jos *tehdä* + *toimenpide* -rakenteen kanssa esiintyy vielä kolmas verbi, on se kaikissa tapauksissa verbi *tehdä*, toisin sanoen sitä ei korvata muilla mahdollisilla verbeillä *ryhtyä*, *suorittaa*, *soveltaa*, *säätää* tai *toteuttaa*.

Taulukko 21. Verbin *tehdä* esiintymien prosenttiosuudet positioittain yhteisesiintymissä *tehdä* + *toimenpide*.

L5	L4	L3	L2	L1	toimenpide, EU	R1	R2	R3	R4	R5
0,0	31,6	0,0	31,6	5,2	toimenpide	0,0	0,0	0,0	0,0	31,6
L5	L4	L3	L2	L1	toimenpide, SL	R1	R2	R3	R4	R5
2,9	1,5	13,2	27,9	2,9	toimenpide	41,2	8,9	0,0	1,5	0,0

Esimerkissä 23 yhteisesiintyjä *tehdä* sijaitsee positioissa L4, L2 ja R5 (EU-aineisto) ja esimerkissä 24 se on positioissa L3 ja R1 SL-aineisto).

- (23) mukaisesti tekemiä tai tehtäväksi annettuja toimenpiteitä pidetään kolmanteen nähden tekeminä (EU)
- (24) uhalla, että tekemättä jätetty toimenpide teetetään laiminlyöjän kustannuksella tai että ||Rivi 2907244 (SL)

SL-aineistossa esiintymiä on kaikissa muissa positiossa lukuunottamatta positioita R3 ja R5. Eniten esiintymiä on positiossa R1, josta löytyy 27 esiintymää, mikä on noin prosentti kaikista hakusanan *toimenpide* esiintymistä verrannaisaineistossa.

Yhteisesiintymä *tehdä* + *toimenpide* on selvästi produktiivisempi SL-aineistossa kuin EU-aineistossa. Verbi *tehdä* on myös tuttu yhteisesiintyjänä muissa ammatti-

kielissä, esimerkiksi lääketieteen erikoiskielessä, jossa lääketieteellisiä toimenpiteitä intuitiivisen arvioni ja Kotuksen verkkosanakirjan perusteella tyypillisesti *tehdään* (Kotuksen sanakirjan hakutulos löytyy kuvasta 7.1 luvusta 7.)

f) *toteuttaa + toimenpide*

EU-aineistossa yhteisesiintymä *toteuttaa + toimenpide* esiintyy 656 kertaa. Verbi *toteuttaa* on siten selkeästi tavallisin saneen *toimenpide* yhteisesiintyjä. Peräti 30 prosenttia eli lähes joka kolmas *toimenpide* esiintyy aineistossa yhdessä verbin *toteuttaa* kanssa (kun verbi on tutkimusikkunan alueella) ja koko aineistossa se esiintyy normalisoituna yli 54 kertaa per 100 000 sanetta. SL-aineistossa esiintymiä on yhteensä vain 90 kappaletta, mikä tarkoittaa hiukan yli kolmea prosenttia hakusanan *toimenpide* kaikista esiintymistä aineistossa ja normalisoituna 2,5 esiintymää per 100 000 sanetta. (X-squared = 63.753, df = 1, p-value = 1.41e-15, joten voidaan todeta hyvin pieni p-arvo (< 0.05) -> Verbin *toteuttaa* suhteelliset osuudet tutkimuskohteen *toimenpide* tutkitussa tekstiyhteydessä poikkeavat aineistoissa toisistaan, ero on tilastollisesti merkitsevä ja riippuvuusoletus voidaan hylätä).

EU-aineistossa yhteisesiintymässä *toteuttaa + toimenpide* verbi on useimmiten positiossa L2 (204 kappaletta, noin 10 prosenttia hakusanan esiintymistä ja lähes 18 prosenttia yhteisesiintymistä *toteuttaa + toimenpide*). Seuraavaksi eniten verbin esiintymiä on positiossa L1 (173 kappaletta, lähes 8 prosenttia hakusanan esiintymistä ja reilu 15 prosenttia yhteisesiintymistä *toteuttaa + toimenpide*). Positiossa L3 on 73 ja positiossa R1 vastaavasti 75 esiintymää (reilu kolme prosenttia ja hiukan alle 4 prosenttia hakusanan esiintymistä). Muissa positioissa esiintymiä on vähemmän. Tarkemmat luvut on esitetty taulukossa 13 ja prosentuaaliset jakaumat löytyvät taulukosta 22.

Taulukko 22. Verbin *toteuttaa* esiintymien prosenttiosuudet positioittain yhteisesiintymissä *toteuttaa + toimenpide*. (EU ja SL)

L5	L4	L3	L2	L1	toimenpide, EU	R1	R2	R3	R4	R5
5,5	5,0	11,1	31,1	26,4	toimenpide	13,1	3,4	1,4	1,8	1,2
L5	L4	L3	L2	L1	toimenpide, SL	R1	R2	R3	R4	R5
2,2	0,0	4,4	26,7	16,7	toimenpide	35,6	5,6	4,4	4,4	0,0

Ensimmäisestä verrannaisaineistosta (SL) eniten esiintymiä löytyy EU-aineiston tavoin positioista R1 (24 kappaletta) ja L2 (32 kappaletta). Luvut vastaavat noin yhtä prosenttia kaikista hakusanan esiintymistä. Positiossa R1 on yhteisesiintymistä

toteuttaa + toimenpide reilu 2 prosenttia ja positiossa L1 prosentin verran. Seuraavaksi eniten niitä on positiossa L1 (15 kappaletta, reilu 0,5 prosenttia kaikista yhteisesiintymistä, joissa mikä tahansa tutkimusverbeistä esiintyy hakusanan *toimenpide* kanssa ja reilu prosentti yhteisesiintymistä *toteuttaa + toimenpide*).

Tutkimusverbi *toteuttaa* sijaitsee EU-aineistosta tyypillisimmin ennen hakusanaa, kun taas SL-aineistossa sen paikka on tavallisimmin hakusanan jälkeen.

EU-aineistossa tyypillisin rakenne on *toteuttaa + ADJ + toimenpide* (positio L2), ja tyypillisin adjektiivi tässä rakenteessa on *tarvittava*, kuten esimerkin 25 positiossa L1. Muita kyseisessä rakenteessa usein esiintyviä adjektiiveja ovat *aiheellinen*, *asianmukainen* ja *tarpeellinen*. *Toteuttaa* positiossa L3 kuvautuu useimmiten rakenteeseen *toteuttaa + kaikki + ADJ + toimenpide*.

- (25) Asianomaisten jäsenvaltioiden on toteutettava tarvittavat toimenpiteet rajoittaakseen ja tarvittaessa arvioidakseen kuhunkin (EU)

EU-aineistossa tyypillisiä yhteisesiintyjä positioittain: *toteuttaa + toimenpide*

Toimenpide on EU-aineistossa yksikkömuotoinen 33 tapauksessa 2180 esiintymästä. Esimerkin 26 kanssa identtisiä rakenteita on 173 kpl. Yhteisesiintyjä *toteuttaa* on positiossa L1.

- (26) keinot tai niiden on toteutettava toimenpiteet sen varmistamiseksi, että vastaavilla (EU)

Esimerkin 27 kanssa identtisiä rakenteita on 87 kpl. Yhteisesiintyjä on niissä positiossa L2. Adjektiivin vaihtelua rakenteessa: *aiheellinen* (20), *asianmukainen* (12), *tarpeellinen* (25), *tällainen* (4), *tarkoituksenmukainen* (3). Verbiyhteisesiintyjä *toteuttaa* on positiossa L2 yhteensä 204 kappaletta. (EU)

- (27) Toimivaltaisen viranomaisen on toteutettava tarvittavat toimenpiteet antaakseen jokaiselle hyväksytyin rakennetyypin mukaan (EU)

Esimerkin 28 kanssa identtisiä rakenteita on 25 kpl. Yhteisesiintyjä on tuolloin positiossa L3. Adjektiivin vaihtelua rakenteessa: *aiheellinen* (8), *asianmukainen* (4), *tarpeellinen* (18), *tällainen* (4), *tarkoituksenmukainen* (3). Verbiyhteisesiintyjä *toteuttaa* on positiossa L3 yhteensä 74 kappaletta (EU-aineisto).

Esimerkin 29 kanssa identtisiä rakenteita EU-aineistossa on 5 kpl. Yhteisesiintyjä on tällöin positiossa L4. Adjektiivin vaihtelua rakenteessa: *tarpeellinen* (2), *tällainen*

- (28) niiden on toteutettava kaikki tarvittavat toimenpiteet näiden seuraamusten täytäntöönpanon varmistamiseksi (EU)

(4), *tarkoituksenmukainen* (3). Verbiyhteisesiintyjä *toteuttaa* löytyy positiosta L4 yhteensä 33 kappaletta.

- (29) kuultuaan toteutettava tätä varten tarpeelliset toimenpiteet 1 päivään heinäkuuta 1961 mennessä (EU)

Esimerkin kanssa identtisiä rakenteita EU-aineistossa on 5 kpl. Yhteisesiintyjä on tällöin positiossa L5. Adjektiivin vaihtelua rakenteessa: *tarpeellinen* (2), *suhteellinen* (2), *tarkoituksenmukainen* (3). Verbiyhteisesiintyjä *toteuttaa* löytyy positiosta L5 yhteensä 36 kappaletta. (EU)

- (30) kuultuaan toteutettava kaikki tätä varten tarpeelliset toimenpiteet 1 päivään heinäkuuta 1961 mennessä (EU)

Esimerkin 31 kanssa identtisiä rakenteita on 75 kpl (sisältää verbikantaisen substantiivin *toteuttaminen*), yhteisesiintyjä on tällöin positiossa R1 (EU).

- (31) säädetyt ylärajat, jos nämä toimenpiteet toteutetaan perustamissopimuksen 92, 93 (EU)

Esimerkissä 32 yhteisesiintyjä on EU-aineistossa positiossa R2. Esimerkin kanssa identtisiä rakenteita on 10 kpl ja esimerkin 33 kanssa identtisiä rakenteita on 2 kpl. Yhteisesiintyjä on tuolloin positiossa R3. Esimerkeissä 34 ja 35 yhteisesiintyjä on positiossa R4. Jälkimmäinen on myös esimerkki konkordanssista, jossa positiossa R1 on välimerkki. Näitä on EU-aineistossa 9 kappaletta. Esimerkissä 36 yhteisesiintyjä on EU-aineistossa positiossa R5. Esimerkin kanssa identtisiä rakenteita EU-aineistossa on 5 kpl.

- (32) kuljeteta pois paikalta ja tarkoituksenmukaiset toimenpiteet on toteutettu yhteisön kasvinjalostajan oikeuden hakijan. (EU)

- (33) että niiden vaikutusten kumoamiseksi tarvittavat toimenpiteet on tarvittaessa toteutettu, sanotun (EU)

- (34) 5 kohdassa tarkoitettu säädös tai toimenpide on annettu taikka toteutettu, (EU)

- (35) interventiot; c) Yhteiset toimenpiteet, jotka on toteutettu perustamissopimuksen (EU)
- (36) interventiot; c) käsiteltäviksi, ja b) toimenpiteet, jotka jäsenvaltioiden on toteutettava. (EU)

EU-aineistossa tyypillisen verbin *toteuttaa* esiintymisestä SL-aineistossa

SL-aineistossa *toimenpide* esiintyy noin kolmasosassa (28/90) konkordansseista +/-5 saneen tutkimusikkunassa. Nämä lainkohdat liittyvät Eurooppalaiseen yhteistyöhön, ja tekijänä niissä on yleensä jäsenvaltio, kuten alla olevassa esimerkissä 37:

- (37) jäsenvaltioiden on toteutettava tarvittavat toimenpiteet sen varmistamiseksi, että myyjä (SL)

Aineiston valossa EU-asetusten esiintymä *toimenpide* merkitsee sellaista asiaa, joka ensisijaisesti *toteutetaan*. Erityisesti toteutetaan *aiheelliset, asianmukaiset, tarpeelliset, tarvittavat ja välttämättömät toimenpiteet* päätösten ja säädösten noudattamiseksi. Euroopan unionin aihepiirejä vastaavissa teksteissä SL-aineistossa ei ennen vuotta 1998 ole juuri *toteutettutoimenpiteitä* muutoin kuin eurooppalaisen yhteistyön nimissä.

Adjektiivin vaihtelua yhteisesiintymän *toteuttaa + toimenpide* yhteydessä

Vakiorakenteet, vakiofraasit ja vakiolausekkeet ovat yksi säädöskielen ominaispiirteistä. Niiden käyttö on osoittaa säädöskielen tuntemusta ja sille tyypillisen selkeyden ja johdonmukaisuuden ymmärtämistä. Olen siksi kiinnittänyt tässä tutkimuksessa huomiota usein toistuvien rakenteiden, fraasien ja lausekkeiden esiintymiseen sekä tutkimuskohteena olevassa EU-asetusten aineistossa että sen verrannaisaineistossa (SL). Tarkemmat tiedot aineistoista löytyvät luvusta 4.

Yhteisesiintymä *toimenpide + toteuttaa* muodostaa adjektiivien kanssa aineistossa usein toistuvan vakiorakenteen *toteuttaa + kaikki + ADJ + toimenpide*. Esitän rakenteesta esimerkin 38.

- (38) Jäsenvaltioiden on toteutettava kaikki aiheelliset toimenpiteet sen varmistamiseksi, että omien

Adjektiivien esiintymien lukumäärät positiossa L1 esimerkin rakenteessa ovat seuraavat: *tarvittava* 25, *tarpeellinen* 18, *asianmukainen* 4 ja *aiheellinen* 7, kun ky-

seistä rakennetta on EU-aineistosta 54 kappaletta. Adjektiivien *tarvittava* ja *tarpeellinen* frekvenssit tässä yhteisesiintymässä *toteuttaa toimenpide* viittaa kahteen kilpailevaan vakiolausekkeeseen.

6.1.3 Yhdyssana yhteisesiintyjänä

Tutkimusaineistossa ja ensimmäisessä verrannaisaineistossa kaikki yhdyssanat, joiden yhtenä osana on *toimenpide*, ovat määrityyhdyssanoja. Ne ovat substantiiveja, lukuun ottamatta EU-aineistossa esiintyvää moniosaista adjektiivista yhdyssanaa *avustustoimenpidekohtainen*, joita aineistossa on kaksi kappaletta. Lekseemi *toimenpide* esiintyy useammin yhdyssanan loppuosana EU-aineistossa kuin SL-aineistossa (585 kappaletta 2 793 esiintymää kohden EU-aineistossa ja 445 kappaletta 3 253 esiintymää kohden SL-aineistossa). EU-aineistossa 43 esiintyy vain kerran, kun taas SL-aineistossa vain kerran esiintyviä vastaavia yhdyssanoja on 37 kappaletta. Molemmista aineistoista löytyviä yhdyssanoja, joissa *toimenpide* on edusosana, löytyy 12 kappaletta. Taulukossa 23 esitän lukumäärät lekseemin *toimenpide* esiintymisestä yhdyssanan osana.

Taulukko 23. Saneen *toimenpide* esiintyminen tutkimusaineistoissa yksinään, yhdyssanan loppuosana, yhdyssanan alkuosana sekä yhdyssanan keskellä.

Tyyppi = sananmuototyyppi; Esiintymä = sananmuotoesiintymä.

<i>toimenpide</i>	EU tyyppi	EU esiintymät	SL tyyppi	SL esiintymät	Yhteisiä
yksinään		1	2180	1	2763
yhdyssanan loppuosana	89	585	87	445	12
yhdyssanan alkuosana	4	22	14	44	
yhdyssanan keskellä	3	6	1	1	
yhteensä	97	2793	103	3253	

EU-aineistossa tavallisin näistä *toimenpide*-yhdyssanoista on *interventioitoimenpide* (60 esiintymää) ja muita lähes yhtä tavallisia yhdyssanoja ovat *tasoitustoimenpide* (59 esiintymää) ja *valvontatoimenpide* (52 esiintymää) ja *erityistoimenpide* (52 esiintymää). SL-aineistossa on vain yksi yhdyssana, joka esiintyy yli 50 kertaa: *työtaiistelutoimenpide* (62 esiintymää). *Interventioitoimenpiteen* lisäksi kahden muun EU-aineiston yhdyssanan määriteosa on käännöslainan englannin kielestä: *suspensioitoimenpide* (2 esiintymää) ja *tariffitoimenpide* (9 esiintymää). Kaikki ovat kuitenkin jo vakiintuneet suomen kieleen; *suspensio* ja *tariffi* löytyvät itsenäisinä hakusanoina Kotuksen verkkosanakirjasta (tarkistettu tammikuussa 2016).

Määritelyyhdyssanojen kollokaatioprofilien voidaan olettaa noudattavan samantyyppistä profilia lekseemin *suojatoimenpide* kanssa. Tarkastelen seuraavaksi tilannetta

kahden yhdyssanan osalta, jotka ovat *suojatoimenpide* ja *valvontatoimenpide* (taulukko 24). Valintaperusteena on esiintymien lukumäärä molemmissa aineistoissa. Molemmissa aineistoissa esiintyviä yhdyssanoja, joissa *toimenpide* on viimeisenä osana, on vain 12 kappaletta, mikä rajoittaa tutkittavien yhdyssanojen valintaa. Voidaan siis todeta, että EU-aineistossa ja SL-aineistossa on erilaisia yhdyssanoja.

Taulukko 24. Tutkittavien verbien *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä* ja *toteuttaa* esiintyminen yhdyssanojen *suojatoimenpide* ja *valvontatoimenpide* kanssa (EU) ja (SL)

Hakusana	EU hakusanat	yhteisesiintyjä	EU kpl	positiot
suojatoimenpide	32	<i>toteuttaa</i>	13	L1: 6, R1: 3, R2: 4
valvontatoimenpide	52	<i>soveltaa</i>	7	L5: 1, R1: 6
		<i>suorittaa</i>	1	R3: 1
		<i>säätää</i>	8	L5: 1, L1: 7
		<i>toteuttaa</i>	3	L3: 1, L2: 1, R5: 1
	SL hakusanat	yhteisesiintyjä	SL kpl	positiot
suojatoimenpide	9	<i>ryhtyä</i>	2	L1: 1, R2: 1
		<i>soveltaa</i>	1	R1: 1
valvontatoimenpide	15	<i>säätää</i>	1	L1: 1
		<i>tehdä</i>	2	R2: 1, R3: 1

Varsinaisessa tutkimusaineistossa (EU) tutkittujen yhdyssanojen *suojatoimenpide* ja *valvontatoimenpide* yhteisesiintyjät tutkittujen verbien joukossa painottuvat muutoinkin aineistossa hakusanan *toimenpide* kanssa usein esiintyviin verbeihin, etupäässä verbiin *toteuttaa*. Se on ainoa tutkituista verbeistä, joka esiintyy yhdyssanan *suojatoimenpide* kanssa, mutta se on vasta kolmanneksi tavallisin yhteisesiintyjä yhdyssanan *valvontatoimenpide* tekstiyhteydessä, jossa tutkituista verbeistä tavallisin on *säätää*. Yhdyssanan *suojatoimenpide* profiilit EU-aineistossa ja SL-aineistossa eroavat toisistaan. Yhdyssanalla *valvontatoimenpide* on niissä sama verbikollokaati yhdessä tapauksessa (verbi *säätää*).

interventiotoimenpide

Seuraavaksi tarkastelen tarkemmin yhdyssanan *interventiotoimenpide* kanssa esiintyviä verbejä. Yhdyssana valikoitui tutkimuskohteeksi osaltaan sen vuoksi, että ai-

neistosta löytyi yhdyssana *interventiotoiminta* ja *toiminta* oli aluksi ehdokkaana tutkittavaksi substantiiviksi. *Interventiotoimenpide* edustaa lisäksi EU-sanaston myötä Suomen säädöskieleen saapunutta uudissanaa, jota ei löydy verrannaisaineistosta (SL). Arvelen sen yhteisesiintyjien selventävän kuvaa tutkimuskohteesta *toimenpide*. Yhdyssana *interventiotoimenpide* esiintyy EU-aineistossa kaikkiaan 60 kertaa ja 23 tapauksessa sen tekstiyhteydestä löytyy jokin kuudesta tutkittavasta verbistä (taulukko 25).

Taulukko 25. Hakusanan *interventiotoimenpide* tunnusluvut määrällisten tulosten vertailussa. (EU ja SL)

	EU	SL
1) koko aineiston saneiden lukumäärä	1371479	3561992
1b) normalisointikerroin	0,073	0,028
2) kaikki saneen <i>interventiotoimenpide</i> esiintymät aineistossa	60	0
2b) normalisoituna per 100 000 esiintymää; esiintymistääjuus	4,4	0
3) kaikkien tutkimuskohteena olevien kuuden verbin kaikki yhteisesiintymät hakusanan <i>interventiotoimenpide</i> kanssa	23	0
3b) normalisoituna per 100 000 esiintymää	1,7	0

Esitän tutkittavien kuuden verbin frekvenssit hakusanan *interventiotoimenpide* tekstiyhteydessä taulukossa 26.

Yhdyssanan *interventiotoimenpide* kollokaatioprofiili suhteessa tutkittaviin verbeihin poikkeaa lekseemin *toimenpide* profilista seuraavasti: verbi *säätää* esiintyy 20 prosentissa kaikista yhdyssanan *interventiotoimenpide* esiintymistä, kun taas vastaava prosenttiluku hakusanan *toimenpide* tekstiyhteydessä on noin 9. *Toteuttaa* (15 prosenttia) puolestaan ei ole yhtä taajaan esiintyvä yhdyssanan yhteisesiintyjänä kuin itsenäisen lekseemin yhteisesiintyjänä, jolloin verbin *toteuttaa* esiintymistääjuus on yli 29 prosenttia. Myös verbi *soveltaa* esiintyy yhdyssanan yhteisesiintyjänä noin puolessa saneen *toimenpide* esiintymistä itsenäisenä (3,3 prosenttia vs. 8,5 prosenttia).

Havaintojen perusteella voidaan todeta, että EU-aineistossa mainitut yhdyssanat noudattavat suunnilleen hakusanan *toimenpide* kaavaa, toisin sanoen SL-aineistoissa tyypillisiä verbikollokaatteja *ryhtyä* tai *suorittaa* ei hakutuloksista löydy eikä myös-

Taulukko 26. Yhdyssanan *interventiotoimenpide* kanssa yhteisesiintymissä esiintyvien verbien kappalemäärät ja prosenttiosuudet kaikista hakusanan *interventiotoimenpide* esiintymistä positiivittain tutkimusaineistossa. (EU)

EU											
Interventiotoimenpide	kpl	L5	L4	L3	L2	L1	R1	R2	R3	R4	R5
<i>soveltaa</i>	2						2				
<i>säätää</i>	12		3		9						
<i>toteuttaa</i>	9	2				6	1				
YHTEENSÄ	23										
	%										
<i>soveltaa</i>	3,3						3,3				
<i>säätää</i>	2		5		15						
<i>toteuttaa</i>	15	3,3				10	1,7				
YHTEENSÄ	38,3										

kään kummassakin aineistossa suhteellisen vähän esiintyvää verbiä *tehdä*. Sen sijaan edustettuina ovat verbit *soveltaa*, *säätää* ja *toteuttaa*.

Verbin *säätää* sisältävistä konkordansseista 9 on vakiorakenteen tyyppiä *säätää* + *erityinen* + *interventiotoimenpide*, konkordanssi on esitetty esimerkissä 39.

- (39) 1418/76 6 artiklassa säädetyt erityiset interventiotoimenpiteet .
3 Asetuksen (ETY || Rivi 183724 (EU)

interventio yhteisesiintyjänä

Yhdyssanan *interventiotoimenpide* määriteosa *interventio* esiintyy EU-aineistossa itsenäisenä lekseeminä 99 kertaa. Se muodostaa yhdyssanoja myös muiden substantiivien kanssa, jotka on esitetty alla olevassa taulukossa 27.

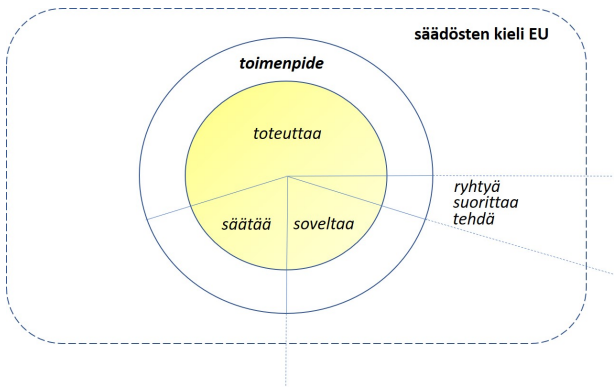
6.1.4 Yhteisesiintyjien jakaumat

Molemmissa aineistossa EU ja SL noin puolessa termin *toimenpide* esiintymistä sen yhteisesiintyjänä on jokin tutkimusverbeistä, mutta niiden jakaumat eroavat toisistaan. Esitän jakaumat kuvissa 6.1 ja 6.2.

Varsinaisessa tutkimusaineistossa (EU) termin *toimenpide* kanssa tavallisin yhteisesiintyjä tutkimusverbien joukossa on *toteuttaa*, jonka osuus yhteisesiintymistä on yli puolet. Verbit *säätää* ja *soveltaa* ovat suunnilleen yhtä yleisiä yhteisesiintyjä, kun taas verbien *ryhtyä*, *suorittaa* ja *tehdä* osuus yhteisesiintyjistä on yhteensä alle

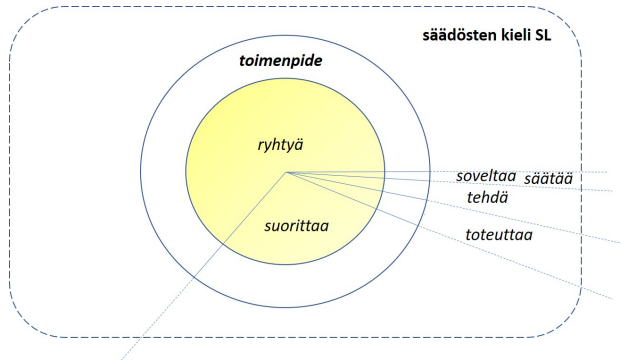
Taulukko 27. *interventio*-alkuiset yhdyssanat. (EU)

interventio	99	julkinen
interventio		
-elin	122	
-toimielin	73	
-tuote	17	
-osto	3	
-kustannus	3	
-meno	2	
-hintaa	13	perusinterventiohintaa (2)
-makasiini	1	
-varasto	39	
-varastointi	4	
-väline	1	
-voi	4	erityismyynti (8 esiintymää, joista 3 <i>interventiovoimaa</i>)



Kuva 6.1: Hakusanan *toimenpide* kanssa esiintyvien tutkimusverbiin jakauma EU-aineistossa ja niiden osuus tutkimusikkunassa. Hakusanan esiintymiä on 2280 ja tutkimusverbit esiintyvät sen kanssa 1142 tapauksessa.

kahdeksasosa. Verbiin jakaumat yhteisesiintyjienä ensimmäisessä verrannaisainestossa (SL) ovat lähes päinvastaiset: verbiin *ryhtyä* osuus on lähes kaksi kolmasosaa, verbiin *suorittaa* noin neljäsosa ja verbit *toteuttaa* ja *tehdä* esiintyvät yhteensä vajaa kahdeksasosassa yhteisesiintyjienä ja verbiin *soveltaa* ja *säätää* osuus on alle kymmenesosa.



Kuva 6.2: Hakusanan *toimenpide* kanssa esiintyvien tutkimusverbien jakauma SL-aineistossa ja niiden osuus tutkimusikkunassa. Hakusanan esiintymiä on 2762 ja tutkimusverbit esiintyvät sen kanssa 1450 tapauksessa.

6.2 Muutos säädösteksteissä (semfinlex-lait)

Tässä osiossa tarkastelen tutkimuskohteiden määrällistä esiintymistä toisessa verrannaisaineistossa (semfinlex-lait), jonka tarkempi kuvaus löytyy aineistoluvusta 4. Tarkastelen tutkimuskohteen *toimenpide* esiintymistä aineistossa ja sen osakorpuksissa itsenäisenä lekseiminä sekä osana tutkittuja yhteisesiintymiä. Tutkittava ajanjakso kattaa Suomen lainsäädännössä ennen varsinaisen tutkimusaineiston (EU) ja sen ensimmäisen verrannaisaineiston (SL) kokoamista sekä sen jälkeen kolmen ajanjakson sisällä. Esiintymät on kerätty eri aikakausilta, jotta voidaan arvioida, onko tutkittavan termin *toimenpide* yhteisesiintymän esiintymistajuus ja sen esiintyminen osana yhteisesiintymää tutkittavien verbien kanssa ajan mittaan muuttunut. Kielipankin konkordanssityökalu Korpin käyttö ja aineistoon koodatut tekstiattribuutit mahdollistavat tutkimusikkunan +/- 5 ulkopuolisten havaintojen tekemisen, koska rajaukset voidaan tehdä jo hakuvaiheessa.

Kielipankissa julkaistu laaja Semfinlex-aineisto sisältää osakorpuksen, jossa on Finlexin alkuperäisiä säädöksiä vuosilta 1734–2018 ja oletettavasti myös ensimmäisen verrannaisaineiston (SL) aineksia, vaikka SL-aineisto kattaakin vain osan lainsäädännön aihealueista. Pällekkäisyyksiä ei kuitenkaan ole pystytty tämän tutkimuksen puitteissa tarkistamaan. Haut tutkimuskohteen *toimenpide* yhteisesiintymistä tutkimusverbien kanssa Kielipankin konkordanssityökalu Korpissa on tehty kolmelta ajanjaksolta seuraavasti:

Ensimmäisen tulosjoukon ajallinen rajausta sisältää vuodet 1960–1999. Tutkimuksen ensimmäinen verrannaisaineisto on koottu vuoden 1999 teksteistä, minkä vuoksi ajanjakso on valittu päätymään jo vuonna 1999. Tämän osion tutkimuksen tavoit-

teenä on tuottaa tietoa siitä, miten taaajaan lakseemiä *toimenpide* on käytetty ennen EU-asetusten kääntämistä ja sisällyttämistä Suomen lainsäädäntöön 2000-luvun alussa.

Toinen haku on rajattu sisältämään ajanjakson 1999–2009 ja kolmas ajanjakson 2010–2018. Tarkastelemalla saneen *toimenpide* esiintymistä eri jaksoissa saadaan tietoa sen käytön yleistymisestä sekä sen kanssa yhteisesiintyvien verbien lukumäärien kehittymisestä ajan mittaan. Kaikkien hakujen tekstilaji on laki.

Sane *toimenpide* esiintyy aineiston lakiteksteissä vain 171 kertaa ajanjaksolla, joka päättyy vuonna 1960, toisin sanoen ennen tähän tutkimukseen valittujen tekstien (EU) tuottamisen ajanjaksoa. Esiintymien jakaantuminen eri vuosikymmenille on esitetty taulukossa 28 alla.

Taulukko 28. Hakusanan *toimenpide* esiintyminen ennen tutkimuksessa tarkemmin tarkasteltuja ajanjaksoja. (semfinlex-lait)

ajanjakso	<i>toimenpide</i>	saneiden määrä	prosenttiosuus
1734–1889	0	22 482	0
1890–1899	4	4 910	0,08
1900–1909	0	2 070	0
1910–1919	10	5 709	0,17
1920–1929	14	36 780	0,04
1930–1939	27	29 884	0,09
1940–1949	15	59 219	0,02
1950–1959	101	124 414	0,08
yhteensä	171	285 468	

Semfinlex-aineiston sisältämät lait kattavat koko lainsäädännön kentän, kun taas SL-aineisto on koottu siten, että se sisältää aihealueet, jotka ovat EU-aineiston kanssa mahdollisimman yhtäpitäviä. Näin SL-aineisto vastaa mahdollisimman tarkasti EU-aineiston tilannekontekstia. Semfinlex tarjoaa näin ollen mahdollisuuden tutkimusaineiston (EU) ja sen verrannaisaineiston (SL) määrällisistä analyyseistä saatujen tulosten vertaamiseen edellä mainittujen kolmen eri ajanjakson aikana voimassa olleeseen laajempaan Suomen lainsäädäntöön (alkuperäisiin säädöksiin), rajattuna säädöskokoelman sisältämiin lakeihin. On huomattava, että säädöskokoelmasta on ajan mittaan korvattu lainsäädäntöä, joten esitetyt lukumäärät eivät toteuta diakronisen tutkimuksen edellytyksiä. Sen sijaan ne ovat ikkuna tämänhetkisen (2018) säädöskokoelman sisältämiin vanhempiin säädösteksteihin.

Saneen *toimenpide* esiintymisessä aineiston lakiteksteissä ennen tutkimukseen

sisältyviä ajanjaksoja ei ole havaittavissa suuria eroja. Termin käyttö on lisääntynyt suhteessa tekstimassan kasvamiseen, vaikkakin vuosina 1910–1919 ja 1940–1949 havaintojen lukumäärä on pienempi kuin muina ajanjaksoina.

Koko tekstilajin laki sisältävän osakorpuksen saneiden lukumäärä on 12 833 980. Tästä on otettu tarkasteluun lait 1.1.1960 alkaen: ajanjakson 1.1.1960–31.12.2018 sisällä lait sisältävät yhteensä 11 523 595 sanetta. Ne jakaantuvat eri ajanjaksoille seuraavasti: vuosina 1960–1998 esiintymiä on yhteensä 3 822 946 sanetta, vuosina 1999–2009 yhteensä 3 690 320 sanetta ja vuosina 2010–2018 yhteensä 4 010 329 sanetta. Semfinlex-aineiston lakiteksteissä on 28 041 saneen *toimenpide* esiintymää. EU-aineistossa esiintymiä on 2 180 ja koko aineistossa saneita on yhteensä 1 371 479.

Toimenpide on tavallisempi sane EU-aineistossa kuin semfinlex-lait-aineistossa. Ensimmäinen verrannaisaineisto (SL) on koottu vuonna 1999, joten sen sisältämiä lakeja saattaa sisältyä semfinlex-lait-aineiston ensimmäisen ajanjakson 1960–1998 tuloksiin.

Peilaan seuraavassa varsinaisesta tutkimusaineistosta (EU) ja ensimmäisestä verrannaisaineistosta (SL) saatuja määrällisiä tuloksia semfinlex-lait-osakorpuksen, milloin tutkimustulokset antavat siihen aiheita. Myös joitakin positiokohtaisia havaintoja ja vakiolausekkeita on käsitelty alla jokaisen tutkimusverbin osalta erikseen. Verbin esiintymistä yhteisesiintymän osana EU-aineistossa ja semfinlex-lait-aineistossa verrataan toisiinsa Khiin neliö -testin avulla. Aineisto on Kielipankin konkordanssityökalussa (Korp) haun yhteydessä rajattu kattamaan eri aikakausia siten, että haut on tehty hakuikkunassa hakusana +/- 5, toisin sanoen L5–R5, seuraavilta ajanjaksoilta: 1) 1960–1998, 2) 1999–2009 ja 3) 2010–2018. Lisäksi on tehty ylimääräinen haku hakuikkunassa L15–L6 + R6–R15 kattamaan hakuikkunan ulkopuolella olevia tutkimusverbin esiintymiä.

Tässä osiossa esitän ensin yhteisesiintymien lukumäärät eri ajanjaksoilta taulukoissa verbikohtaisesti. Sen jälkeen esitän ristiintaulukon, josta selviävät yhteisesiintymien tilastollisessa laskennassa käytetyt muuttujien arvot. Muuttujat (lukumäärät) ovat seuraavat: 1) verbi esiintyy yhteisesiintymän osana ja 2) verbi esiintyy muualla aineistossa (mahdollisesti muissa yhteisesiintymissä). Aineistot, joita verrataan keskenään ovat EU ja semfinlex-lait. Tilastollisen testin tuloksen jälkeen esitän havaintojen kehittymisen ajanjaksoina 1960–1998, 1999–2009 ja 2010–2018 graafisessa muodossa. Tilastolliset analyysit on tehty varsinaisen tutkimusaineiston (EU) kokoamisen jälkeiseltä ajanjaksolta 1999–2018.

a) *ryhtyä* + *toimenpide*

Verbin *ryhtyä* esiintymiä aineistossa semfinlex-lait on 8 095 kappaletta. Näistä yli viidesosassa se esiintyy osana yhteisesiintymää *toimenpiteen* kanssa.

Verbi *ryhtyä* esiintyy hakusanan *toimenpide* yhteisesiintyjänä kaikkina tutkituina ajanjaksoina (ks. taulukko 29), ennen vuotta 1999 vähän useammin kuin sen jälkeen. Se ei kuitenkaan ole juurikaan rakentamassa yhteisesiintymiä, joissa yhteisesiintyjä on hakuikkunan +/- 5 sanetta ulkopuolella. Painotus verbin sijoittumisessa hakusanan jälkeen indikoi verbikantaisia substantiiveja tai passiivirakenteita.

Taulukko 29. Tutkimusverbin *ryhtyä* esiintyminen hakusanan *toimenpide* kanssa samassa rakenteessa tutkimusikkunassa. (semfinlex-lait)

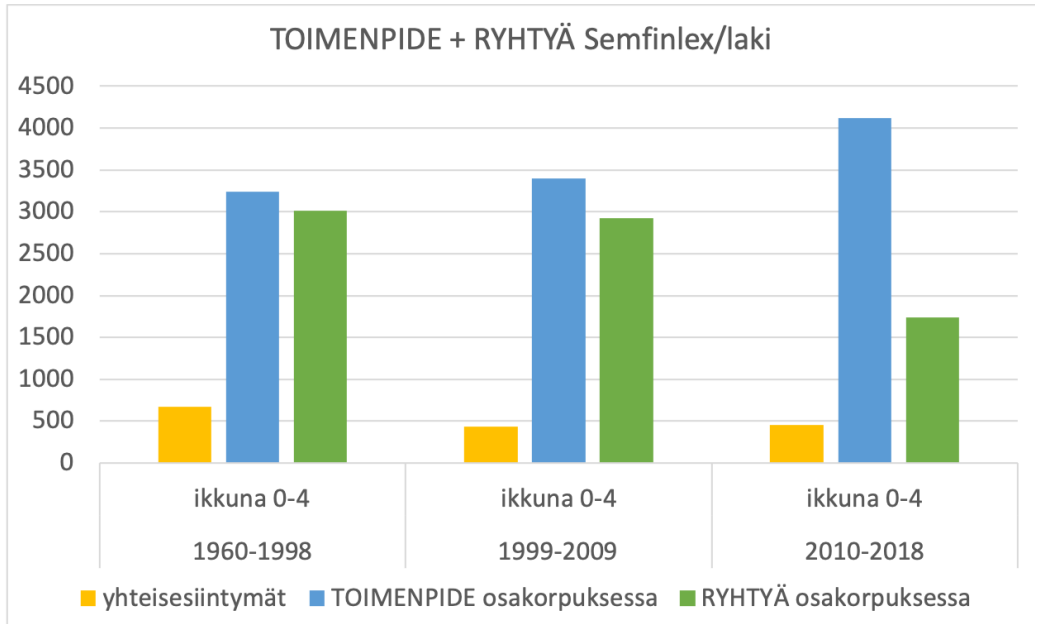
ajanjakso	<i>toimenpide</i> L1–L5	<i>toimenpide</i> R1–R5	yhteensä
1960–1998	207	467	674
1999–2009	124	306	430
2010–2018	127	330	457
yhteensä	458	1103	1561
ajanjakso	<i>toimenpide</i> L6–L15	<i>toimenpide</i> R6–R15	yhteensä
2010–2018	31	21	52

Taulukko 30. Ristiintaulukko: yhteisesiintyjän *ryhtyä* esiintyminen yhteisesiintymässä *ryhtyä* + *toimenpide*. (EU ja semfinlex-lait)

yhteisesiintymän komponenttina	EU	semfinlex-lait 1999–2018
<i>ryhtyä</i> esiintyy	40	887
<i>ryhtyä</i> ei esiinny	80	3779

Testin tuloksena on niin pieni merkitsevyysarvo, että voidaan hylätä nollahypoteesi, jonka mukaan verbin *ryhtyä* suhteellinen frekvenssi yhteisesiintymässä *ryhtyä* + *toimenpide* on sama molemmissa aineistoissa EU ja semfinlex-lait. Havaittujen suhteellisten frekvenssien ero aineistojen välillä on tilastollisesti merkitsevä. (X-squared = 15.369, df = 1, p-value = 8.843e-05; Fisherin kaksipuoleinen eksakti testi: p-value = 0.0002418) Toteamme, että EU-aineistossa on vain vähän verbin *ryhtyä* esiintymiä ja niistäkin puolet esiintyy yhteisesiintymässä *toimenpiteen* kanssa.

Tulos vahvistaa varsinaisen tutkimusaineiston (EU) ja ensimmäisen verrannaisai-



Kuva 6.3: Tutkimusverbin *ryhtyä* esiintyminen osana yhteisesiintymää *ryhtyä + toimenpide*. (semfinlex-lait)

neiston (SL) havaintoja yhteisesiintymistä *ryhtyä + toimenpide* ajanjaksolta 1960–1998. Verbin rooli yhteisesiintyjänä on kuitenkin pienentynyt vuoden 1998 jälkeen. Koska saneen *toimenpide* lukumäärät säädöskielessä ovat vuoden 1998 jälkeen kasvaneet tasaisesti, voidaan olettaa, että verbin *ryhtyä* tilalle on tullut ja tulossa muita yhteisesiintyjä.

Yksityiskohtainen positioiden analyysi ei ole ollut mahdollista verbistä *ryhtyä* hakusanan *toimenpide* yhteisesiintyjänä aineiston laajuuden vuoksi. Esiintymistajuksien perusteella huomataan, että sekä verbin *ryhtyä* että yhteisesiintymän *ryhtyä + toimenpide* lukumäärät aineistossa vähenevät vuoden 1998 jälkeen (kuva 6.3 ja taulukko 29). Verbi *ryhtyä* on siitä huolimatta edelleen tavallisin saneen *toimenpide* yhteisesiintyjä aineistossa. Voidaan kuitenkin todeta, että pääosa havainnoista on positioissa R1–R5, joissa havaintoja on 1 103, kun taas lukumäärä positioissa L1–L5 on 458. Tutkimusikkunan ulkopuolelta tehdyt havainnot painottuvat kuitenkin hakusanaa ennen sijaitseviin positioihin L6–L15, joissa on 31 havaintoa, kun taas positioissa R6–R15 esiintymiä on 21.

b) *soveltaa* + *toimenpide*

Verbin *soveltaa* esiintymiä aineistossa semfinlex-lait on 36 581 kappaletta.

Verbiä *soveltaa* ei käytetä termin *toimenpide* kanssa semfinlex-lait-aineistossa osana tutkittua yhteisesiintymää oikeastaan lainkaan. Tutkimustulos on tässä suhteessa selkeä. Yhteisesiintymä *soveltaa* + *toimenpide* on vieläkin harvinaisempi kaikkina tutkittuina ajanjaksoina kuin SL-aineistossa. Tulokset on esitetty taulukossa 31.

Taulukko 31. Tutkimusverbin *soveltaa* esiintyminen hakusanan *toimenpide* kanssa samassa rakenteessa tutkimusikkunassa. (semfinlex-lait)

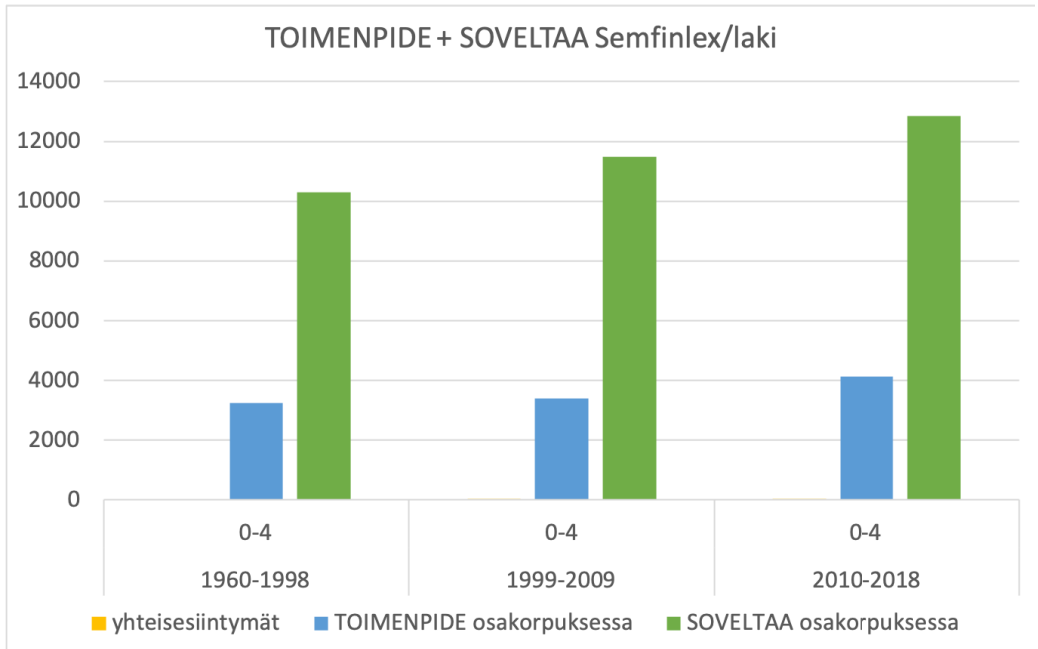
ajanjakso	<i>toimenpide</i> L1-L5	<i>toimenpide</i> R1-R5	yhteensä
1960–1998	0	0	0
1999–2009	2	0	2
2010–2018	2	2	4
yhteensä	4	2	6
ajanjakso	<i>toimenpide</i> L6-L15	<i>toimenpide</i> R6-R15	yhteensä
2010–2018	0	4	4

Taulukko 32. Ristiintaulukko: yhteisesiintyjän *soveltaa* esiintyminen yhteisesiintymässä *soveltaa*+ *toimenpide*. (EU ja semfinlex-lait)

yhteisesiintymän komponenttina	EU	semfinlex-lait 1999–2018
<i>soveltaa</i> esiintyy	198	6
<i>soveltaa</i> ei esiinny	5 038	24 313

Testin tulokseksi saadaan niin pieni merkitsevyysarvo, että voidaan hylätä nolalahypoteesi, jonka mukaan verbin *soveltaa* suhteellinen frekvenssi yhteisesiintymässä *soveltaa* + *toimenpide* on sama molemmissa aineistoissa EU ja semfinlex-lait. Havaittujen suhteellisten frekvenssien ero aineistojen välillä on tilastollisesti merkitsevä. (Khiin neliö -testi: X-squared = 899.67, df = 1, p-value < 2.2e-16; Fisherin kaksipuoleinen eksakti testi: p-value < 2.2e-16)

Vaikka *soveltaa* on tavallinen verbi semfinlex-lait-aineistossa ja sen käyttö säädöskielessä on lisääntynyt ajan myötä (kuva 6.4), yhteisesiintyjänä *toimenpiteen* kanssa sitä ei juuri esiinny. Tulos vahvistaa havaintoa *soveltaa*-verbin epätyypill-



Kuva 6.4: Tutkimusverbin *soveltaa* esiintyminen osana yhteisesiintymää *soveltaa + toimenpide*. (semfinlex-lait)

sestä käytöstä *toimenpiteen* yhteydessä EU-asetuksissa.

Tutkimusikkunan L5–R5 ulkopuolella on yhtä paljon verbin *soveltaa* esiintymiä kuin sisäpuolellakin, neljä molemmissa. Esitän tästä esimerkin 40, jossa verbi *soveltaa* sijaitsee ikkunan L5–R5 ulkopuolella positiossa R9 (semfinlex-lait).

- (40) sovellettava tietosuojalain (1050/2018) 6 §:n 2 momentissa säädettyjä toimenpiteitä rekisteröidyn oikeuksien

c) *suorittaa + toimenpide*

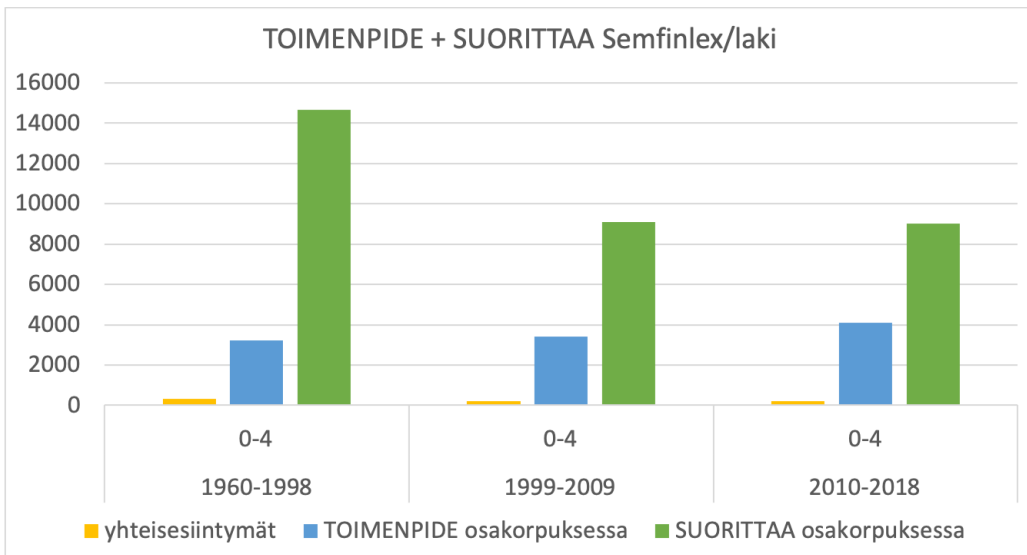
Verbin *suorittaa* esiintymiä aineistossa semfinlex-lait on 37 136.

Verbi *suorittaa* on verbin *soveltaa* vastakohta: se on huomattavasti tavallisempi *toimenpiteen* yhteisesiintyjä SL-aineistossa verrattuna EU-aineistoon. Semfinlex-aineiston tulokset löytyvät taulukosta 33.

Testin tulokseksi saadaan merkitsevyysarvo, joka on niin suuri, että Khiin neliö-testin perusteella ei voida hylätä nollahypoteesia, jonka mukaan verbin *suorittaa*

Taulukko 33. Tutkimusverbin *suorittaa* esiintyminen hakusanan *toimenpide* kanssa samassa rakenteessa tutkimusikkunassa. (semfinlex-lait)

ajanjakso	<i>toimenpide</i> L1-L5	<i>toimenpide</i> R1-R5	yhteensä
1960–1998	188	147	335
1999–2009	122	79	201
2010–2018	121	93	214
yhteensä	431	319	750
ajanjakso	<i>toimenpide</i> L6–L15	<i>toimenpide</i> R6–R15	yhteensä
2010–2018	0	0	0



Kuva 6.5: Tutkimusverbin *suorittaa* esiintyminen osana yhteisesiintymää *suorittaa* + *toimenpide*. (semfinlex-lait)

suhteellinen frekvenssi yhteisesiintymässä *suorittaa* + *toimenpide* on sama molemmissa aineistoissa EU ja semfinlex-lait. Havaittujen suhteellisten frekvenssien ero aineistojen välillä ei ole tilastollisesti merkitsevää. (Khiin neliö -testi: X-squared = 8.2531, df = 1, p-value = 0.004068; Fisherin kaksipuoleinen eksakti testi: p-value = 0.003231)

Verbin *suorittaa* lukumäärä osana tutkittavaa yhteisesiintymää *suorittaa* + *toimenpide* toisessa verrannaisaineistossa semfinlex-lait on laskenut vuoden 1998 jäl-

Taulukko 34. Ristiintaulukko: yhteisesiintyjän *suorittaa* esiintyminen yhteisesiintymässä *suorittaa + toimenpide*. (EU ja semfinlex-lait)

yhteisesiintymän komponenttina	EU	semfinlex-lait 1999–2018
<i>suorittaa</i> esiintyy	23	415
<i>suorittaa</i> ei esiinny	1 807	17 714

keisen kymmenvuotiskauden aikana ja pysynyt sen jälkeen suunnilleen samalla tasolla (kuva 6.5). Myös yhteisesiintymien *suorittaa + toimenpide* lukumäärät ovat samalla pienentyneet noin kolmanneksella. Tulos vahvistaa varsinaisen tutkimusaineiston (EU) ja ensimmäisen verrannaisaineiston (SL) yhteisesiintymän *suorittaa + toimenpide* tarkastelusta tehtyä johtopäätöstä verbin *suorittaa* aliedustuksesta EU-aineistossa verrattuna SL-aineistoon.

Verbi *suorittaa* esiintyy hiukan enemmän hakusanaa ennen sijaitsevista positioista L1-L5 kuin sen jälkeisissä positioissa R1-R6. Yhtään esiintymää ei löydy tutkimusikkunan rajojen ulkopuolelta.

Esimerkissä 41 verbit *suorittaa* ja *tehdä* sijaitsevat positioissa L4 ja L2 (semfinlex-lait).

- (41) suorittaa tai teettää tarvittavat toimenpiteet tienpitäjän kustannuksella

d) *säätää + toimenpide*

Verbin *säätää* esiintymiä aineistossa semfinlex-lait on 98 073 kappaletta.

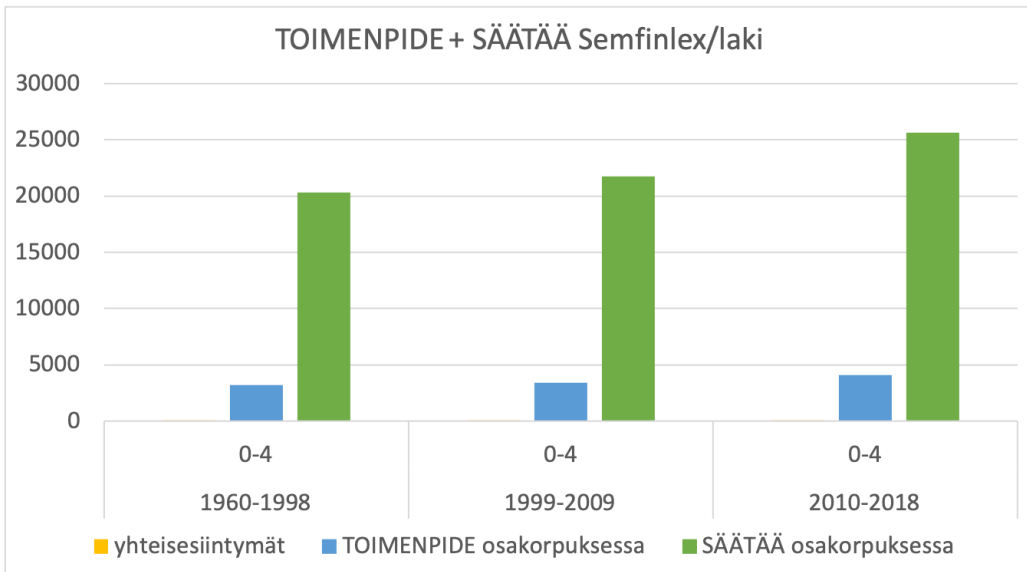
Verbi *säätää* on aineistossa semfinlex-lait lähes yhtä harvinainen yhteisesiintymässä *säätää + toimenpide* kuin *soveltaa*, kuten taulukosta 35 käy ilmi. Se on vielä harvinaisempi kuin SL-aineistossa, jossa yhteisesiintymiä on 17.

Vaikka *säätää* on luonnollisesti tavallinen verbi semfinlex-lait-aineistossa ja sen määrä on ajan mittaan tasaisesti kasvanut (kuva 6.6), se on yhteisesiintyjänä *toimenpiteen* kanssa niin harvinainen, että positiokohtainen tarkastelu ei ole mielekäs. Tulos vahvistaa havaintoa *säätää*-verbin epätyypillisestä käytöstä *toimenpiteen* yhteydessä EU-asetuksissa.

Testin tuloksena on niin pieni merkitsevyysarvo, että sen perusteella voidaan hylätä nollahypoteesi, jonka mukaan verbin *säätää* suhteellinen frekvenssi yhteisesiin-

Taulukko 35. Tutkimusverbin *säätää* esiintyminen hakusanan *toimenpide* kanssa samassa rakenteessa tutkimusikkunassa. (semfinlex-lait)

ajanjakso	<i>toimenpide</i> L1-L5	<i>toimenpide</i> R1-R5	yhteensä
1960–1998	2	0	2
1999–2009	2	0	2
2010–2018	1	0	1
yhteensä	5	0	5
ajanjakso	<i>toimenpide</i> L6-L15	<i>toimenpide</i> R6-R15	yhteensä
2010–2018	0	0	0



Kuva 6.6: Tutkimusverbin *säätää* esiintyminen osana yhteisesiintymää *säätää + toimenpide*. (semfinlex-lait)

tymässä *säätää + toimenpide* on sama molemmissa aineistoissa EU ja semfinlex-lait. Havaittujen suhteellisten frekvenssien ero aineistojen välillä on tilastollisesti merkitsevä. (Khiin neliö -testi: X-squared = 2348.8, df = 1, p-value < 2.2e-16; Fisherin kaksipuoleinen eksakti testi: p-value < 2.2e-16)

Taulukko 36. Ristiintaulukko: yhteisesiintyjän *säätää* esiintyminen yhteisesiintymässä *säätää*+ *toimenpide*. (EU ja semfinlex-lait)

yhteisesiintymän komponenttina	EU	semfinlex-lait 1999–2018
<i>säätää</i> esiintyy	203	3
<i>säätää</i> ei esiinny	3 838	47 374

e) *tehdä* + *toimenpide*

Verbin *tehdä* esiintymiä aineistossa semfinlex-lait on 42 731 kappaletta.

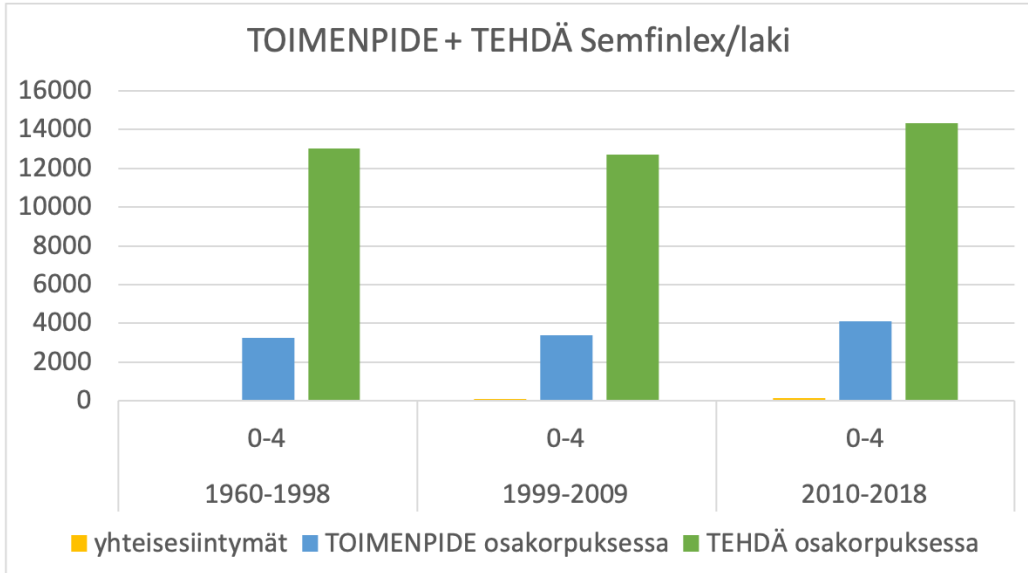
Verbi *tehdä* on ainoa tutkimusverbien joukossa, joka esiintyy osana yhteisesiintymää *tehdä* + *toimenpide* sekä EU-aineistossa (19 esiintymää) että SL-aineistossa (68 esiintymää). EU-aineistossa kyseessä on tosin identtinen virke ja 19 esiintymää löytyvät 7 konkordanssiriviltä. Semfinlex-aineistossa *tehdä* esiintyy yhdessä saneen *toimenpide* kanssa EU-aineiston ja SL-aineiston ajanjaksolla 66 kertaa ja se on muita tutkimusverbejä (lukuunottamatta verbiä *säätää*) harvinaisempi yhteisesiintyjä (taulukko 37).

Testin tuloksena on niin pieni merkitsevyysarvo, että voidaan hylätä nollahypoteesi, jonka mukaan verbin *tehdä* suhteellinen frekvenssi yhteisesiintymässä *tehdä* + *toimenpide* on sama molemmissa aineistoissa EU ja semfinlex-lait. Havaittujen suhteellisten frekvenssien ero aineistojen välillä on tilastollisesti merkitsevä. (X-squared = 12.082, df = 1, p-value = 0.0005091)

Taulukko 37. Tutkimusverbin *tehdä* esiintyminen hakusanan *toimenpide* kanssa samassa rakenteessa tutkimusikkunassa. (semfinlex-lait)

ajanjakso	<i>toimenpide</i> L1–L5	<i>toimenpide</i> R1–R5	yhteensä
1960–1998	12	54	66
1999–2009	21	80	101
2010–2018	42	110	152
yhteensä	75	244	319
ajanjakso	<i>toimenpide</i> L6–L15	<i>toimenpide</i> R6–R15	yhteensä
2010–2018	4	20	24

Verbi *tehdä* on kaikkina ajanjaksoina ollut tavallinen verbi toisessa verrannaisaineistossa ja sen esiintyminen komponenttina yhteisesiintymässä *tehdä* + *toimenpide* on kasvanut ajanjaksolla 2010–2018 (kuva 6.7).



Kuva 6.7: Tutkimusverbin *tehdä* esiintyminen osana yhteisesiintymää *tehdä + toimenpide*. (semfinlex-lait)

Taulukko 38. Ristiintaulukko: yhteisesiintyjän *tehdä* esiintyminen yhteisesiintymässä *tehdä + toimenpide*. (EU ja semfinlex-lait)

yhteisesiintymän komponenttina	EU	semfinlex-lait 1999–2018
<i>tehdä</i> esiintyy	19	253
<i>tehdä</i> ei esiinny	4 506	26 777

Testin tuloksena on niin pieni merkitsevyysarvo, että voidaan hylätä nollahypoteesi, jonka mukaan verbin *tehdä* suhteellinen frekvenssi yhteisesiintymässä *tehdä + toimenpide* on sama molemmissa aineistoissa EU ja semfinlex-lait. Havaittujen suhteellisten frekvenssien ero aineistojen välillä on tilastollisesti merkitsevä. (X-squared = 12.082, df = 1, p-value = 0.0005091)

Aineistossa ajanjaksolta 2010–2018 tavallisin adjektiivin rakenteessa V + ADJ + N on *tarpeellinen* ja tavallisin vakiorakenne on *tekemättä jätetty toimenpide teetätään asianomaisen / kustannuksella / suostumuksella*. Samoin kuin SL-aineistossa, suurin osa yhteisesiintyjän havainnoista on positioissa L2 ja R1. Tutkimusikkunan ulkopuolelta on havaintoja positioista R6–R5, joissa esiintymiä on 20 kappaletta ja positioissa L6–15 neljä kappaletta.

f) *toteuttaa* + *toimenpide*

Verbin *toteuttaa* esiintymiä aineistossa semfinlex-lait on 6 761 kappaletta.

Verbin *toimenpide* esiintyminen *toimenpiteen* yhteisesiintyjänä semfinlex-lait-aineistossa on lisääntynyt ajan mittaan (ks. taulukko 39). Osa-aineistojen saneiden lukumäärät ovat myös kasvaneet, mutta eivät merkittävästi (1960–1998 yht. 3 822 946 sanetta, 1999–2009 yht. 3 690 320 sanetta ja 2010–2018 yht. 4 010 329 sanetta).

Taulukko 39. Tutkimusverbin *toteuttaa* esiintyminen hakusanan *toimenpide* kanssa samassa rakenteessa tutkimusikkunassa. (EU ja semfinlex-lait)

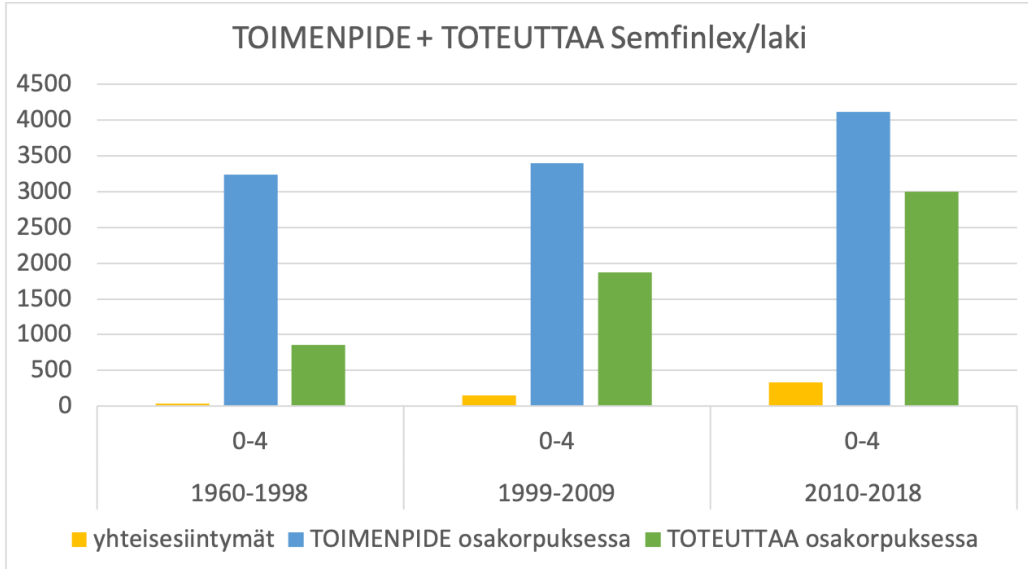
ajanjakso	<i>toimenpide</i> L1–L5	<i>toimenpide</i> R1–R5	yhteensä
1960-1998	27	5	32
1999-2009	102	46	148
2010-2018	158	183	341
yhteensä	287	234	521
ajanjakso	<i>toimenpide</i> L6–L15	<i>toimenpide</i> R6–R15	yhteensä
2010-2018	25	27	52

Verbin *toteuttaa* lukumäärä aineistossa on kaksinkertaistunut kahden ensimmäisen vuosikymmenen kuluessa ja vuosina 2010–2018 sen käyttö on lisääntynyt yli kolmanneksella verrattuna edelliseen vuosikymmeneen 1999–2009 (kuva 6.8). Myös verbin osuus *toimenpiteen* yhteisesiintyjänä on kasvanut lähes viisinkertaiseksi vuosina 1999–2009 verrattuna edelliseen vuosikymmeneen.

Verbin *toteuttaa* lukumäärä aineistossa on kaksinkertaistunut kahden ensimmäisen vuosikymmenen kuluessa ja vuosina 2010–2018 sen käyttö on lisääntynyt yli kolmanneksella verrattuna edelliseen vuosikymmeneen 1999–2009 (kuva 6.8). Myös verbin osuus *toimenpiteen* yhteisesiintyjänä on kasvanut lähes viisinkertaiseksi vuosina 1999–2009 verrattuna edelliseen vuosikymmeneen.

Verbi *toteuttaa* on poikkeuksellinen myös sen vuoksi, että se jakautuu muita verbejä tasaisemmin positioihin L5–R5 ja esiintyy myös näiden ulkopuolella positioissa L6–L15 ja R6–R15. Hakusana *toimenpide* saa mahdollisesti enemmän määreitä, kun yhteisesiintymänä on verbi *toteuttaa* (esimerkki 42). Aineistossa 2010–2018 tavallimmat adjektiivit rakenteessa V + ADJ + N ovat *tarpeellinen* ja *tarvittava*.

(42) toteutettava kiireellisesti tilapäinen ennaltaehkäisevä toimenpide , erityisesti



Kuva 6.8: Tutkimusverbin *toteuttaa* esiintyminen osana yhteisesiintymää *toteuttaa* + *toimenpide*. (semfinlex-lait)

Taulukko 40. Ristiintaulukko: yhteisesiintyjän *toteuttaa* esiintyminen yhteisesiintymässä *toteuttaa*+ *toimenpide*. (EU ja semfinlex-lait)

yhteisesiintymän komponenttina	EU	semfinlex-lait 1999–2018
<i>toteuttaa</i> esiintyy	656	477
<i>toteuttaa</i> ei esiinny	748	4396

Testin tuloksena saadaan niin suuri merkitsevyysarvo, että sen perusteella voidaan hylätä nollahypoteesi, jonka mukaan verbin *toteuttaa* suhteellinen frekvenssi yhteisesiintymässä *toteuttaa* + *toimenpide* on sama molemmissa aineistoissa EU ja semfinlex-lait. Havaittujen suhteellisten frekvenssien ero aineistojen välillä on tilastollisesti merkitsevä. (Khiin neliö -testi: X-squared = 2348.8, df = 1, p-value < 2.2e-16; Fisherin kaksipuoleinen eksakti testi: p-value < 2.2e-16)

Positiokohtainen tarkastelu paljastaa, että havaintoja on tasaisesti hakusanan molemmilla puolilla niin positioissa L1–L5 (287 havaintoa) ja R1–R5 (234 havaintoa) kuin positioissa L6–L15 (25 havaintoa) ja R6–R15 (27 havaintoa).

Yhteenveto

Tässä yhteenvedossa kokoan yhteen edellä esitetyt määrälliset havainnot, jotka on tehty toisesta verrannaisaineistosta (semfinlex-lait).

Toisen verrannaisaineiston tehtävänä oli tuottaa vertailukohde varsinaisen tutkimusaineiston (EU) ja ensimmäisen verrannaisaineiston (SL) tuottamille tuloksille. Tulos vahvistaa varsinaisen tutkimusaineiston (EU) ja ensimmäisen verrannaisaineiston (SL) havaintoja yhteisesiintymistä *ryhtyä* + *toimenpide* ajanjaksolta 1960–1998. Verbin rooli yhteisesiintyjänä on kuitenkin vähentynyt vuoden 1998 jälkeen. Koska saneen *toimenpide* lukumäärät säädöskielessä ovat vuoden 1998 jälkeen kasvaneet tasaisesti, voidaan olettaa, että verbin *ryhtyä* tilalle on tullut ja tulossa muita yhteisesiintyjä.

Tulos vahvistaa havaintoja verbien *soveltaa* ja *säätää* epätyypillisestä käytöstä *toimenpiteen* yhteydessä EU-asetuksissa ja verbin *suorittaa* aliedustuksesta yhteisesiintymän *suorittaa* + *toimenpide* komponentteina EU-aineistossa verrattuna SL-aineistoon.

Verbin *tehdä* osalta tutkimustulos tarkoittaa, että sen rooli Suomen säädösteksteissä on arveltua suurempi ja sen esiintyminen yhteisesiintymän *tehdä* + *toimenpide* komponenttina ei poikkea tilastollisesti EU-aineiston tuloksista.

Verbin *toteuttaa* lukumäärä aineistossa on kaksinkertaistunut kahden ensimmäisen vuosikymmenen kuluessa ja vuosina 2010–2018 sen käyttö on lisääntynyt yli kolmanneksella verrattuna edelliseen vuosikymmeneen 1999–2009. Myös verbin osuus *toimenpiteen* yhteisesiintyjänä on kasvanut lähes viisinkertaiseksi vuosina 1999–2009 verrattuna edelliseen vuosikymmeneen.

Toisen verrannaisaineiston yhteisesiintymien tarkastelu suhteessa varsinaisen tutkimusaineistosta (EU) ja ensimmäisestä verrannaisaineistosta (SL) tehtyihin havaintoihin vahvistaa tuloksia verbien *ryhtyä*, *soveltaa*, *suorittaa* ja *säätää* osalta, kun ne esiintyvät osana yhteisesiintymää tutkimuskohteen *toimenpide* kanssa.

6.3 Tekstilajikontaisia havaintoja (LaDi)

Tässä alaluvussa tarkastelen tutkimuskohteiden määrällistä esiintymistä kolmannessa verrannaisaineistossa Lakeja ja Direktiivejä (LaDi), jonka tarkempi kuvaus löytyy aineistoluvusta 4. Alunperin kun LaDi oli ainoa käytettävissäni oleva lakikielen verrannaiskorpus SL-korpuksen lisäksi ennen Semfinlex-aineiston julkaisemista, oli sen tehtävänä tuottaa tietoa siitä, onko yhteisesiintymissä havaittavissa muutosta varsinaisen tutkimusaineiston ajanjaksolta 1980–1999 ja jälkeen vuosilta 1999–2003.

Pidän tutkimustulokset mukana edelleen, kahdesta syystä. Ensimmäinen on aineiston pienuus. Konkordanssit on mahdollista käsitellä ja epäkelvot hakutulokset poistaa manuaalisesti. Toinen syy liittyy tekstilajiin 'direktiivit', joka ei ole tässä tutkimuksessa tutkimuksen kohteena. Direktiivit ovat EU-säädöksiä, jotka suomennetaan asetusten tavoin, mutta jotka Suomessa käyvät läpi valmisteluprosessin kotimaisten säädösten tapaan. Kurkistus direktiiveihin on pieni ylimääräinen lenkki määrällisten havaintojen ketjussa. Kielipankin konkordanssityökalu Korp tarjoaa mahdollisuuden tarkastella verbien esiintymistä LaDi-aineistossa tekstikontekstina koko virke.

Korpuksen tekstilajit jakaantuvat alalajeihin seuraavasti. Direktiivien alalajeja ovat Euroopan parlamentin ja neuvoston direktiivi, Komission direktiivi ja Neuvoston direktiivi. Asetusten alalajeja ovat Euroopan parlamentin ja neuvoston asetukset ja Neuvoston asetukset (tarkemmin alaluvussa 4.3. Näiden lisäksi lait muodostavat oman alalajinsa. LaDi-aineiston tekstejä on siten tuotettu erilaisissa tilanne- ja tekstikonteksteissa ja ne on koottu useista lähteistä. Olen laskenut tutkimuskohteiden esiintymät direktiiveissä ja asetuksissa sisältämät alalajeittain, jotta mahdolliset poikkeamat saadaan näkyviin, mutta yhteenvetoihin alalajien määrälliset tutkimustulokset on yhdistetty. Varsinaisessa tutkimusaineistossa (asetuksia, EU) ja ensimmäisessä verrannaisaineistossa (lakeja, SL) eroa ei ollut tarpeen tehdä, koska kaikki tekstit käännettiin tuolloin keskitetysti.

EU-tekstejä on aineistossa yhteensä 41 ja niistä vain neljä on asetuksia. Lakeja on yli puolet aineiston teksteistä. Jokaisesta tekstilajista olen kirjannut absoluuttiset frekvenssit, mutta ei suhteellisia frekvenssejä ei ole saatavissa, koska vain koko aineiston saneiden lukumäärä on tiedossa. Aineistoa ei myöskään voi Korpin avulla tarkastella osakorpuksina. Prosenttiosuudet on laskettu alalajeittain verbien määristä. Varsinaisessa tutkimusaineistossa prosenttiosuudet on laskettu kyseisen aineiston kaikkien tutkimusverbien esiintymistä, joten prosenttiosuuksia ei voi verrata keskenään. LaDi-aineistossa niiden tehtävä on tarjota kokonaiskuva verbikollokaatin jakautumisesta eri positiioihin.

Ydinkysymys on, tuottaako hiukan uudempi aineisto varsinaiseen tutkimusaineistoon (EU) verrattuna toisenlaisia määrällisiä tuloksia, onko tutkittavien substantiivien kanssa esiintyvien verbien osuus erilainen ja jakautuvatko ne eri tavoin tutkimusikkunan positiioihin.

Tarkentava kysymys 1: Miten paljon tutkimuskohteen *toimenpide* esiintymiä LaDi-aineistossa on, ja miten usein niiden yhteydessä esiintyy jokin verbeistä *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä* tai *toteuttaa*?

Tarkentava kysymys 2: Onko toinen LaDi-aineisto yhteisesiintymien esiintymistajuuksien suhteen samanlaista kuin ensisijainen tutkimusaineisto (EU) ja ensimmäinen verrannaisaineisto (SL)? Lähtötilanne: aineistossa on kolmenlaisia tekstila-

jeja. Suurin osa teksteistä on direktiivejä (37/66), seuraavaksi eniten aineistossa on lakeja (25/66) ja vähiten on asetuksia (4/66). Havainnot on kirjattu absoluuttisina frekvensseinä, joiden tulkinnassa on huomioitava, että kyseisen tekstilajin tai sen alalajin saneiden lukumäärä ei ole ollut tiedossa tätä tutkimusta tehtäessä. Palaan tähän kysymykseen tarkemmin johtopäätösluvussa 8.

Esitän ensin määrälliset havainnot tutkimuskohteen *toimenpide* esiintymisestä LaDi-aineistossa.

Taulukko 41. Hakusanan *toimenpide* esiintymien lukumäärät aineistossa sekä tutkimusverbien *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä* ja *toteuttaa* esiintymät yhteisesiintymissä V + hakusana *toimenpide* yhteensä.

	direktiivit	asetukset	lait
1) koko aineiston saneiden lukumäärä		232449	
2) kaikki saneen <i>toimenpide</i> esiintymät aineistossa	284	63	85
3) kaikkien tutkimuskohteena olevien kuuden verbin kaikki yhteisesiintymät hakusanan <i>toimenpide</i> kanssa	158	15	35

Taulukko 42. Yhteisesiintymän *toteuttaa* + *toimenpide* esiintymien jakautuminen tekstilajeihin **direktiivi**, **asetus** tai **laki** aineistossa LaDi tekstien julkaisuajankohdan mukaan.

Tekstilaji	tekstit ajanjaksolta	esiintymät <i>toteuttaa</i>	yhteensä <i>toteuttaa</i>	<i>toteuttaa</i> + <i>toimenpide</i>	muita verbin <i>toimenpide</i> kanssa esiintyviä substantiiveja
Direktiivi	1980–1999	61			
Direktiivi	2000–2002	161			
			222	143	hanke, järjestelyt, menettely, sisämarkkinat, tehtävä, toimi, toiminto
Asetus	1999	19			
Asetus	2000–2002	22			
			41	12	arviointi, investointi, seuranta, toiminta, toimi
Laki	2000–2003	12			
			12	8	ehto, hanke, järjestelmä

Taulukko 43. Tutkimusverbien *ryhtyä, soveltaa, suorittaa, säätää, tehdä* ja *toteuttaa* esiintymät yhteisesiintymissä tekstilajien alalajeissa. (LaDi)

LaDi		
TOIMENPIDE	esiintymät	normalisoitu/ 100 000
aineiston sanamäärä	232449	
kaikki esiintymät	432	186
esiintymät: direktiivit	284	122
esiintymät: asetukset	63	27
esiintymät: lait	85	37

Verbin *toteuttaa* osuus kaikista tutkituista yhteisesiintymistä, joissa hakusanan *toimenpide* parina on jokin verbeistä *ryhtyä, soveltaa, suorittaa, säätää, tehdä* tai *toteuttaa* on lähes 70 prosenttia kaikista esiintymistä 148 esiintymällä, kun yhteisesiintymiä V + *toimenpide* on direktiiveissä kaikkiaan 158. Verbin *toteuttaa* käyttö direktiiveissä on lähes kolminkertaistunut ajanjaksolla 2000–2002 (23 tekstiä) verrattuna ajanjaksoon 1980–1999 (15 tekstiä), mikä herättää hypoteesin sen käytön lisääntymisestä Suomen säädöskielessä myös direktiivien kautta. Asetusten lukumäärät ovat niin pieniä (19 vuonna 1999 ja 22 vuosina 2000–2002, kaksi tekstiä kummassakin ajanjaksossa) joten verbin käytön lisäämisestä ei voi esittää arveluja. Aineisto sisältää 25 lakitekstiä vuosilta 2000–2003 ja niissä verbi *toteuttaa* esiintyy 12 kertaa, joista 8 yhteisesiintymän osana.

Hakusana *toimenpide* esiintyy koko aineistossa 432 kertaa, joista se on yksikössä 26 tapauksessa. Niistä kaksi on tutkimusverbien kanssa yhteisesiintyviä. Toinen näistä yhteisesiintyy direktiivissä tutkimusverbin *soveltaa* kanssa (esimerkki 43) ja toinen yhteisesiintyy laissa verbiyhteisesiintyjän *tehdä* kanssa (esimerkki 44).

Direktiiveissä on paljon vaihtelua, mutta *toteuttaa* on silti selvästi tavallisin yhteisesiintyjä tutkittavien verbien joukossa. Se esiintyy direktiiveissä 222 kertaa, joista osama yhteisesiintymää *toteuttaa* + *toimenpide* 143 tapauksessa, toisin sanoen lähes 50 prosentissa havainnoista. Kaikki havainnot ovat Parlamentin ja neuvoston direktiiveistä.

Asetuksissa ei ole juuri vaihtelua tutkittujen verbiyhteisesiintyjien välillä; *toteuttaa* on selkeästi tavallisin 12 esiintymällä.

Lakitekstejä on vain muutama, mutta *ryhtyä* on niissä ehdottomasti tavallisin, sen lisäksi jonkin verran on verbien *toteuttaa* ja *suorittaa* esiintymiä ja vain muutama *säätää* ja *tehdä*. Jos tarkastellaan kysymystä verbien näkökulmasta lakiteksteissä,

Taulukko 44. Tutkimusverbien *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä* ja *toteuttaa* esiintymät yhteisesiintymissä V + hakusana *toimenpide*; tekstilajikohtaiset yhteenvedot.

	LX	L5	L4	L3	L2	L1	R1	R2	R3	R4	R5	RX		luokka kpl	tyyppi yht_dir	tyyppi yht_as	tyyppi yht_la
ryhtyä_dir_ne	0	0	0	0	1	0	0	1	0	0	0	0	ryhtyä	3			
ryhtyä_dir_pajane	0	0	0	0	0	0	0	0	0	0	0	0		0	3		
ryhtyä_as_ne	0	0	0	0	1	0	0	0	0	0	0	0		1		1	
ryhtyä_as_pajane	0	0	0	0	0	0	0	0	0	0	0	0		0		0	
ryhtyä_la	1	0	6	0	1	6	2	1	0	0	2	3		22			22
soveltaa_dir_ne	0	0	0	0	0	0	0	0	0	0	0	0	soveltaa	0			
soveltaa_dir_pajane	1	0	0	0	2	0	3	0	0	0	0	0		6	6		
soveltaa_as_ne	0	0	0	0	0	0	0	0	1	0	0	0		1			
soveltaa_as_pajane	0	0	0	0	0	0	0	0	0	0	0	0		0	1		
soveltaa_la	0	0	0	0	0	0	0	0	0	0	0	0		0			0
suorittaa_dir_ne	0	0	1	0	1	0	0	0	0	0	0	0	suorittaa	3			
suorittaa_dir_pajane	0	0	0	0	1	0	0	0	0	0	0	0		1	4		
suorittaa_as_ne	0	0	0	0	0	0	0	0	0	0	0	0		0			
suorittaa_as_pajane	0	0	0	0	0	0	0	0	0	0	0	0		0		0	
suorittaa_la	0	0	1	0	0	1	0	0	0	0	0	0		3			3
säätää_dir_ko	0	0	0	0	0	1	0	0	0	0	0	0		1			
säätää_dir_ne	0	0	0	0	0	1	0	0	0	0	0	0	säätää	1			
säätää_dir_pajane	0	0	0	0	0	0	0	0	0	0	0	0		0	2		
säätää_as_ne	0	0	0	0	0	1	0	0	0	0	0	0		1			
säätää_as_pajane	0	0	0	0	0	0	0	0	0	0	0	0		0		1	
säätää_la	0	0	0	0	0	1	0	0	0	0	0	0		1			1
tehdä_dir_ne	0	0	0	0	0	0	0	0	0	0	0	0	tehdä	0			
tehdä_dir_pajane	0	0	0	0	0	0	0	0	0	0	0	0		0	0		
tehdä_as_ne	0	0	0	0	0	0	0	0	0	0	0	0		0			
tehdä_as_pajane	0	0	0	0	0	0	0	0	0	0	0	0		0		0	
tehdä_la	0	0	0	0	1	0	0	0	0	0	0	0		1			1
toteuttaa_dir_ko	0	3	0	0	2	0	0	0	0	0	0	0	toteuttaa	5			
toteuttaa_dir_ne	3	1	0	6	31	3	3	2	0	1	2	1		53			
toteuttaa_dir_pajane	5	3	3	8	29	29	3	1	0	1	1	2		85	143		
toteuttaa_as_ne	0	2	1	2	1	6	0	0	0	0	0	0		12			
toteuttaa_as_pajane	0	0	0	0	0	0	0	0	0	0	0	0		0		12	
toteuttaa_la	0	1	1	1	1	3	0	0	1	0	0	0		8			8
													yhteensä	208	158	15	35
															dir	aset	lait

- (43) päätöstä 2002/676/EY (radiotaajuuspäätös) tai muuta yhteistä toimenpidettä soveltamalla. (direktiivi)
- (44) tekemättä jätetty toimenpide tehdään (laki)

toimenpiteisiin lähes yksinomaan *ryhdytään*.

Verbien *ryhtyä*, *soveltaa*, *suorittaa* ja *säätää* esiintymiä aineistossa on kaikissa alalajeissa muutamia (2–6 kpl).

Toimenpiteitä ei *tehdä* missään tekstilajissa LaDi-aineistossa. Sen sijaan niitä *toteutetaan* direktiiveissä, asetuksissa ja jonkin verran myös laeissa. Laeissa niiden osuus on yli 60 prosenttia (mutta havaintoja verbistä *toteuttaa* on vain 12, joista 8 yhteisesiintymän osana). Asetuksissa vastaava osuus on reilu 30 prosenttia.

Tutkimusverbi + *toimenpide* positioittain

Havaintoja verbistä *toteuttaa* on lähes kaikista positioista näissä alalajeissa, mikä on poikkeuksellista, myös verrattaessa verbin *toteuttaa* esiintymiä muista LaDi-aineiston alalajeista tai ensimmäisestä verrannaisaineistosta (SL) tehtyihin havaintoihin. Osa direktiiveissä havaituista esiintymistä sijoittuu tutkimusikkunan ulkopuolelle, kun sen kooksi määritellään +/- 5 sanetta hakusanan ympärillä. Olen koonnut tutkimusverbien prosentuaaliset osuudet eri positioissa taulukkoon 45.

Tutkimusikkunan ulkopuolella olevia esiintymiä (LX tai RX) löytyy direktiiveistä, joissa yksi tutkimusverbin *soveltaa* esiintymä sijaitsee positiossa LX.

Direktiiveissä hakusanan *toimenpide* yhteisesiintyjistä verbin *toteuttaa* tutkimusikkunan ulkopuolisista havainnoista 5 on Parlamentin ja neuvoston direktiiveissä ja 4 Neuvoston direktiiveissä. Asetuksissa tutkimusverbi *toteuttaa* löytyy positioista L5–L1. Puolet niistä (6 kappaletta) sijaitsee positiossa L1. Verbi *toteuttaa* sijaitsee tutkimusikkunan ulkopuolella 6 prosentissa yhteisesiintymistä *toteuttaa* + *toimenpide*, toisin sanoen 8 tapauksessa. Nämä tapaukset on ollut mahdollista löytää aineistosta, koska konkordanssityökalu Korp tuottaa hakusanan kontekstiksi kokonaisen virkkeen.

Laeissa tavallisin verbi *ryhtyä* muodostaa vakiolauseita tai pitkiä vakiolausekkeita, koska esiintymiä löytyy myös positioista LX (5 prosenttia tapauksista) ja RX (14 prosenttia tapauksista). Loput esiintymät ovat jakautuneet positioihin L4 ja L1, joissa molemmissa esiintymiä on 27 prosenttia. Joitakin esiintymiä löytyy myös positioista R1 ja R2. Tutkimusverbin *toteuttaa* esiintymiä on 8 kappaletta ja suurin osa niistä on positiossa L1. Tarkemmat tiedot löytyvät taulukosta 45. Esimerkissä 45 yhteisesiintyjä *toteuttaa* sijaitsee positiossa L9 (LaDi-lait).

- (45) Jos kyseinen yritys ei korjaa puutteita 2 kohdassa tarkoitettun määräajan kuluessa, asianomaisen viranomaisen on toteutettava ehtojen ja velvollisuuksien noudattamisen varmistamiseksi tarpeelliset ja oikeasuhteiset toimenpiteet. (laki)

Taulukko 45. Tutkimusverbien *ryhtyä, soveltaa, suorittaa, säätää, tehdä* ja *toteuttaa* esiintymien prosentuaaliset osuudet eri tekstilajien alalajeittain yhteisesiintymissä V + *toimenpide*.

Direktiivit	LX	L5	L4	L3	L2	L1	toimenpide	R1	R2	R3	R4	R5	RX
ryhtyä	0,0	0,0	0,0	0,0	33,3	0,0	toimenpide	33,4	33,3	0,0	0,0	0,0	0,0
soveltaa	16,7	0,0	0,0	0,0	0,0	33,3	toimenpide	50,0	0,0	0,0	0,00	0,0	0,0
suorittaa	0,0	0,0	25,0	25,0	25,0	25,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
säätää	0,0	0,0	0,0	0,0	0,0	100,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
tehdä	0,0	0,0	0,0	0,0	0,0	0,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
toteuttaa	5,6	4,9	2,1	9,7	43,4	22,4	toimenpide	4,2	2,1	0,0	1,4	2,1	2,1
Asetukset	LX	L5	L4	L3	L2	L1	toimenpide	R1	R2	R3	R4	R5	RX
ryhtyä	0,0	0,0	0,0	0,0	0,0	100,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
soveltaa	0,0	0,0	0,0	0,0	0,0	0,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
suorittaa	0,0	0,0	0,0	0,0	0,0	0,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
säätää	0,0	0,0	0,0	0,0	0,0	100,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
tehdä	0,0	0,0	0,0	0,0	0,0	0,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
toteuttaa	0,0	16,7	8,3	16,7	8,3	50,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
Lait	LX	L5	L4	L3	L2	L1	toimenpide	R1	R2	R3	R4	R5	RX
ryhtyä	4,6	0,0	27,3	0,0	4,6	27,3	toimenpide	9,1	4,5	0,0	0,0	9,1	13,5
soveltaa	0,0	0,0	0,0	0,0	0,0	0,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
suorittaa	0,0	0,0	33,3	0,0	0,0	33,4	toimenpide	33,3	0,0	0,0	0,0	0,0	0,0
säätää	0,0	0,0	0,0	0,0	0,0	100,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
tehdä	0,0	0,0	0,0	0,0	100,0	0,0	toimenpide	0,0	0,0	0,0	0,0	0,0	0,0
toteuttaa	0,0	12,5	12,5	12,5	12,5	37,5	toimenpide	0,0	0,0	12,5	0,0	0,0	0,0

Yhteenveto: LaDi

Tarkentavaan kysymykseen, miten paljon tutkimuskohteen esiintymiä LaDi aineistossa vastauksena on, että tutkittujen ajanjaksojen sisällä *toimenpide* esiintyy 432 kertaa, joista yksikössä vain 26 tapauksessa. Yhteisesiintymistä tutkimusverbien kanssa se on yksikössä vain kahdesti. Tutkimus vahvistaa näin EU-aineistosta tehtyä havaintoa monikkumuotoisen lekseemin tyypillisyydestä erikoiskielessä. Toimenpiteen osuus direktiiveissä näyttää kasvavan, kun verrataan ajanjaksoja 1980–1999 ja 2000–2002 toisiinsa mutta asetuksissa vastaavaa kehitystä ei ole havaittavissa ja lakitekstit ovat kaikki myöhäisemmältä jaksolta. Eniten yhteisesiintymiä, joissa *toimenpide* on perusosa, löytyi Euroopan parlamentin ja neuvoston direktiiveistä.

Muita tutkimusverbejä esiintyy direktiiveissä vain 15 yhteisesiintymän komponenttina. Asetuksissa havaintoja tehtiin 12 ja laeissa 8. Tässäkin aineistossa lakien alalajissa tutkituissa yhteisesiintymissä tavallisin verbi on *ryhtyä*. Havaintoja muista verbeistä on liian vähän, jotta niiden esiintymiä voisi verrata varsinaisesta tutkimusaineistosta (EU) tehtyihin havaintoihin.

Tutkimusikkunan ulkopuolella havaitut, aineiston koon huomioiden lukuisat verbikollokaatit viestittävät säädöskielen rakenteesta, joka sallii hyvin monen elementin erottavan yhteisesiintymän osat toisistaan. Se on myös muistutus säädöskielestä erikoiskielenä, joka hyödyntää pitkiä vakiolausekkeita tai vakiolauseita erityisesti EU-teksteissä.

Aineistossa ainoa taaajaan esiintyvä verbi *toteuttaa* havaitaan selvästi useimmin positioista L1 ja L2, tosin 11 esiintymistä löytyy tutkimusikkunan ulkopuolelta. Varsinaisessa tutkimusaineistossa (EU) verbi *soveltaa* esiintyy yhteisesiintyjänä 198 kertaa, joista 111 positiossa R1 ja LaDi-aineistossa *soveltaa* esiintyy vastaavasti direktiiveissä vain 6 kertaa, joista kolme positiossa R1. Tulokset ovat keskenään samansuuntaisia.

Laeissa tavallisin verbi *ryhtyä* direktiiveissä tavallisin verbi *toteuttaa* esiintyvät tutkimusikkunan ulkopuolella muita verbejä useammin, mikä viittaa niiden olevan osa vakiolauseita tai pitkiä vakiolausekkeita, joissa yhteisesiintyjän ja hakusanan välillä on enemmän kuin 5 sanetta.

Aineiston LaDi analyysin tavoitteena oli tarkastella direktiivejä tutkittavien yhteisesiintymien näkökulmasta, koska ne implementoidaan lainsäädäntöön tavallisen lainsäädäntöprosessin kautta. LaDi-aineistossa *toteuttaa* on tavallisin direktiiveissä, erityisesti Neuvoston sekä Parlamentin ja neuvoston direktiiveissä ja produktiivisin verbeistä osana yhteisesiintymää V + *toimenpide*. Hypoteesina voi todeta, että EU-aineiston mukana tulevat konventiot siirtyvät myös direktiivien mukana Suomen säädöskieleen. Aineisto on suppea, mutta tutkimustulos selkeä.

6.4 Havaintoja yleiskielisestä verrannaisaineistosta (lehdet-osa)

Neljäs verrannaisaineisto, osa Kielipankin 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä sisältävästä korpuksesta (lehdet-osa), poikkeaa käsitteeltänsä muista tutkimusaineistosta.

Havaintoja ei ole tutkittu lajittelemalla Korpin konkordanssin tuloksia manuaalisesti taulukkolaskentaohjelmassa. Menetelmän valinta johtuu ensisijaisesti aineistojen suuresta koosta, mutta myös neljännen verrannaisaineiston roolista tutkimuk-

nessa. Tutkimuksessa hyödynnetään konkordanssien sijaan konkordanssityökalu Korpin tuottamaa sanakuva. Kielipankin korpuksesta tutkimuskohteeksi on valittu osio 'Muut lehdet' vastakohtana 'Tiedelehdille', joka sisältää myös lainsäädäntöön liittyviä tekstejä julkaisevan Virallisen lehden. Tehtävänä on tuottaa tietoa tutkimuskohteen esiintymistiheydestä yleiskielisessä aineistossa sekä vastata kysymykseen siitä, millaisten aihepiirien ympärille tutkimusverbit yleiskielessä ryhmittyvät.

Käsittelen ensin määrälliset yhteisesiintymään liittyvät havainnot ja teen sen jälkeen katsauksen verbien tekstiyhteyteen. Aineisto on kuvattu tarkemmin luvussa 4 ja sanakuva luvussa 5.

Lehdet on järjestetty aineistossa aakkosjärjestykseen, joten tässä tutkimuksessa havainnot on tehty ensin valitsemalla hakukohteeksi koko Muiden lehtien osakorpus ja sen jälkeen sama haku on tehty erikseen Virallisesta lehdestä. Tuloksena on tiedot koko aineistosta ja Virallisesta lehdestä, jonka osuus tuloksista raportoidaan erikseen ja havainnot vähennetään koko aineiston tuloksista.

Virallinen lehti poisluettuna, 1900- ja 2000-luvun lehdissä *ryhdytään toimenpiteisiin, sovelletaan lain ja säännösten lisäksi menetelmää, periaatetta ja mallia ja suoritetaan koulutukseen liittyviä tutkintoja. Lisäksi säädetään lakeja ja kelpoisuuksia, toteutetaan unelmia, itseään ja arvojaan sekä tehdään työtä, tutkimusta, valintoja ja päätöksiä.*

Esitän yhteisesiintymien verbikohtaiset lukumäärät taulukossa 46. Tässä haussa Virallisen lehden tuloksia ei sisälly taulukkoon. Tutkimuskohteen *toimenpide* esiintymien lukumäärä aineistossa on 26 890.

Taulukko 46. Tutkimusverbien *ryhtyä, soveltaa, suorittaa, säätää, tehdä* ja *toteuttaa* esiintymät *toimenpiteen* yhteisesiintyjinä positiioittain. (lehdet-osa)

lehdet-osa			
yhteisesiintyjä	L5-L1	R1-R5	yhteensä
ryhtyä	379	158	537
soveltaa	3	4	7
suorittaa	94	73	167
säätää	1	0	1
tehdä	579	294	873
toteuttaa	215	227	442

Yhteisesiintymän V + *toimenpide* tavallisin yhteisesiintyjä yleiskielisessä lehtiaineistossa on verbi *tehdä* 873 havainnon turvin. Seuraavaksi tavallisin verbi *ryhtyä*

537 esiintymällä. Verbi *suorittaa* on jo harvinaisempi (167 havaintoa).

Muut lehdet -aineistossa (lehdet-osa), Virallinen lehti mukaan luettuna, aineiston tavallisin hakusanan *toimenpide* kanssa esiintyvä verbi on *suorittaa*. Sanakuvan perusteella se esiintyy etumääritteen asemassa, jolloin se kuvautuu tekstissä muotoina *suoritettu toimenpide* tai *suoritettava toimenpide*. Verbin *suorittaa* lisäksi tavallinen verbi aineistossa on *tehdä*.

Kun Virallisen lehden tulokset sisältyvät tuloksiin, vaihtuu taajimpaan esiintyvä verbi verbistä *tehdä* verbiksi *suorittaa*. Muutos tapahtuu kohti toisen verrannaisaineiston jakaumaa, jossa verbi *suorittaa* on verbin *ryhtyä* ohella tavallisin verbi yhteisesiintymässä.

Taulukkoon 47 kirjattujen havaintojen perusteella hahmottelen seuraavaksi tutkimusverbien käyttöyhteyksiä yleiskielisessä sanoma- ja aikakauslehtiaineistossa. Virallisessa lehdessä tutkimuskohteen *toimenpide* kontekstissa esiintyy odotetusti samanlaisia aihe-alueita kuin mitä havaittiin aineiston LaDi tarkastelussa. Niihin kuuluvat oikeuskieleen, talouteen ja prosesseihin liittyvät esiintymät, kuten *laki*, *päätös*, *muistutus* ja *nimi*.

Keskityn tässä katsauksessa sanakuvan tuottamiin esiintymiin, jotka eivät ole yhteisiä Virallisen lehden kanssa. Esimerkiksi sanoma- ja aikakauslehdissä tavallisin verbin *ryhtyä* yhteisesiintyjä on *toimenpide*, mutta koska se esiintyy myös Virallisen lehden listalla, sitä ei tässä katsauksessa käsitellä. Verbin *ryhtyä* tekstikontekstista löytyy yksi sanakuvaan mahtunut yleiskielelle tyypillinen sane, *yrittäjä*. Yhteisesiintymäksi tunnistetaan *soveltaa* ja prosessia kuvaavat *menetelmä*, *periaate* ja *malli*. Verbi *suorittaa* liittyy yleiskielisessä aineistossa vahvasti koulutuksen ja talouden aloihin, joita kuvaavat yhteisesiintyjät *opinnot*, *kurssi*, *ammattitutkinto*, *perustutkinto*, *korkeakoulututkinto*, *loppukauppahinta* ja *elatusapu*. Verbi *säätää* on niin vahvasti sidoksissa Viralliseen lehteen, että sen yhteisesiintyjien joukossa ei ole saneita, jotka eivät esiinny Virallisessa lehdessä. Verbi *tehdä* liittyy prosesseihin, yhteisesiintyjät ovat *työ*, *yhteistyö*, *tutkimus*, *valinta* ja *muistutus*. Lopuksi verbi *toteuttaa* näyttäytyy yleiskielisessä aineistossa Korpin sanakuvan perusteella ihmisen henkiin kasvuun liittyvänä verbinä, ja sen yhteisesiintymät tässä roolissa lehtiaineistossa sisältävät saneet *itse*, *arvo* ja *unelma*.

Katsauksen tavoitteena on tarkastella tutkittavaa yhteisesiintymää ja tutkimusverbejä sen komponentteina yleiskielisessä tekstissä sen selvittämiseksi, millaisissa konteksteissa verbit esiintyvät yleiskielisessä tekstissä. Havainnot selventävät kollokaation kontekstuaalista luonnetta erikoiskielessä. Huomataan, että lehtiaineiston yhteisesiintyjinä verbit *ryhtyä*, *soveltaa* ja *säätää* noudattavat yhdenmukaista määrällistä jakaumaa toisen verrannaisaineiston (SL) kanssa, kun taas verbin *toteuttaa* esiintymiä yhteisesiintymän komponenttina on lehtiaineistossa enemmän kuin

Taulukko 47. Tavallisimmat aineiston lehdet-osa tekstiyhteydessä esiintyvät substantiivit tutkimusverbeittäin sekä havainnot Virallisen lehden osuuksista esiintymistä.

<i>ryhtyä</i>	LEHDET-OSA (13416)	VIRALLINEN LEHTI (560; 4%)
toimenpide	676	232 (34%)
yrittäjä	311	
<i>soveltaa</i>	LEHDET-OSA(12903)	VIRALLINEN LEHTI (1767; 14%)
laki	426	238 (56%)
säännös	251	113 (45%)
menetelmä	220	
periaate	132	
malli	108	
asetus	61	55 (90%)
<i>suorittaa</i>	LEHDET-OSA (39542)	VIRALLINEN LEHTI (15313; 39%)
tutkinto	1 700	221 (13%)
elatusapu	772	
opinnot	473	
kurssi	369	
ammattitutkinto	329	
perustutkinto	237	
loppukauppahinta	119	119 (100%)
korkeakoulututkinto	291	
<i>säätää</i>	LEHDET-OSA (23903)	VIRALLINEN LEHTI (14505; 61%)
nimi	780	780 (100%)
laki	508	120 (24%)
kelpoisuus	172	11 (14%?)
taito	168	168 (100%)
<i>tehdä</i>	LEHDET-OSA (231243/350509)	VIRALLINEN LEHTI (23387; 10%)
työ	14 104	
päätös	5 006	1 067 (21%)
yhteistyö	3 783	
tutkimus	1 751	
valinta	1 280	
muistutus	2 004	2 004 (100%)
ilmoitus	1 807	732 (96%)
muutos	1 104	954 (86%)
<i>toteuttaa</i>	LEHDET-OSA (10310)	VIRALLINEN LEHTI (1469; 14%)
itse	261	
arvo	165	
unelma	133	
hanke	128	25 (19%)

SL-aineistossa (1 442 kappaletta). Aineisto muistuttaa tässä suhteessa enemmän EU-aineistoa. Verbin *tehdä* esiintymiä on tutkimusverbeistä selkeästi eniten sekä aineistossa (yli 350 000 esiintymää) ja tutkitun yhteisesiintymän V + *toimenpide* komponenttina (873 esiintymää).

Seuraava alaluku sisältää yhteisesiintymän rakenteeseen liittyvät havainnot. Todettakoon, että laskennalliset havainnot rakenteeseen liittyvistä ominaispiirteistä on pyritty esittämään määrällisten havaintojen puolella.

6.5 Rakenteeseen liittyvät havainnot

Alaluvussa 2.2.2 totesin tässä tutkimuksessa relevanteiksi rakenteeseen liittyviksi kysymyksiksi seuraavat: 1) minkälaisista elementeistä kollokaatio koostuu, 2) miten kaukana toisistaan sen komponentit ovat ja 3) miten monta komponenttia kollokaatio voi sisältää. Muut tässä rakennetta koskevassa yhteenvedossa esitetyt havainnot ovat syntyneet tutkimuksen aikana.

Ensimmäiseen kysymykseen liittyy tarkastelu yhdyssanasta ensimmäisen asteen yhteisesiintymänä tai vakiintumana, toinen kysymys saa vastauksia positioiden tarkastelun kautta ja kolmas vakiolausekkeista ja vakiolauseista tehdyistä havainnoista. Näiden teknisten kysymysten lisäksi tarkastelen verbikantaisia substantiiveja ja kieliopillista lukua. Koska tutkimuksen lähtökohta on määrällinen tarkastelu, olen pyrkinyt raporttoimaan myös rakenteen ominaispiirteiden määrälliset tulokset alaluvussa 6.1.

6.5.1 Yhdyssana yhteisesiintymänä

Yhdyssana on kiteytynyt niin pitkälle, että se kirjoitetaan suomen kielessä yhteen ja se yhteisesiintyy edelleen muiden kielellisten elementtien kanssa. Yhdyssananmuodostus voidaan nähdä uuden merkitysyksikön tuottamisena, jolloin se tuottaisi merkittävän vakiintuman. Jos yhdyssana säilyy kompositionaalisenä, saattaa yhteen kirjoittaminen olla seuraus vaihtelusta, jossa lekseemit voidaan pitkäänkin kirjoittaa joko erikseen tai yhteen, riippuen kielen puhujan omaksumasta käsityksestä yhteisesiintymän vakiintumisen asteesta. Molempien varianttien tai parafrasain ja yhdyssanan esiintyminen kielessä voi olla aktiivinen valinta ja määrääntyä kielen käyttäjän muiden kielellisten valintojen perusteella, esimerkiksi määritteiden lukumäärän tai syntaktisten rajoitteiden.

Yhteisesiintymisen tuloksena tapahtuva yhdyssanojen muodostuminen tuo Suomen säädöskieleen uutta ainesta. Yhdyssananmuodostuksella voidaan myös ratkaista hankaluuksia, jotka johtuvat jo valmiiksi suomeksi löytyvien termien puutteesta. Osa

yhdyssanoista on ei-kompositionaalisia ja sumeita (sumea kuvaa tässä läpinäkyvän vastakohtaa), mikä voi johtua siitä, että yleisesti ottaen EU:n toimintamallit ja sen säätelyn kohteet yhteiskunnassa ovat kansalaisille tuntemattomia. EU-asiantuntijalle termi *interventiovoi* (ks. tarkemmin alla) saattaa olla hyvin läpinäkyvä ja selkeä termi.

Yhteisesiintyjä rajaa yhdyssanan edusosan merkityskenttää kattamaan kontextinsa sallimia tulkintoja. Kuten Mäkisalo on suomen yhdyssanoja koskevassa tutkimuksessaan osoittanut, millään yhdyssanojen tunnistamiseen vaikuttavista tekijöistä (semanttinen läpinäkyvyys, muodollinen vastaavuus ja osien järjestys) ei ole yksinään ratkaisevaa merkitystä yhdyssanan tunnistamisessa, vaan tunnistamisen prosessiin vaikuttaa useita samanaikaisia tekijöitä (Mäkisalo (2000)). Koska EU-kontekstissa on usein kyse suomalaisille uusista käsitteistä, ovat ne ennen termin Suomen säädöskielen integroitumista suhteellisen tuntemattomia, joten semanttisen läpinäkyvyyden edellyttämä esiintymistajuus ei nouse ratkaisevaksi tekijäksi. Useamman kuin kaksiosaisen yhdyssanan viimeinen osa ei välttämättä myöskään aina ole sen edusosa. Siten ei ole perusteita yksiselitteisesti rajata yhdyssanan edusosa kantamaan pääasiallista merkitystä tai rajaamaan yhdyssanan merkityskenttää, jota määriteosa tarkentaisi. Toisin sanoen EU-aineiston yhdyssanojen jälkiosa ei aina välttämättä ole semanttinen edusosa ja alkuosa tai alkuosat semanttisia määriteosia (ks. johtopäätökset *interventio*-alkuisten yhdyssanojen analyysistä alla).

Yhtä käsitettä kohden voi olla kaksi vaihtoehdoisen synonyymistä termiä, esimerkiksi *myyntitoimi* ja *myyntitoimenpide*. SL-aineistossa ei ole yhtään kummankaan vaihtoehdon esiintymää.

Tutkimuskohde *toimenpide* yhdyssanan edusosana

EU-aineistossa on suhteellisesti enemmän yhdyssanoja kuin SL-aineistossa, kun tarkasteltavana ovat yhdyssanat, joiden yhtenä osana on tutkimuksen kohteena oleva *toimenpide* (ks. yhteenvetotaulukko 48). Tutkimuskohteena oleva substantiivi voi tällöin sijaita missä kohtaa yhdyssanaa tahansa, sen edusosana tai määriteosana tai keskellä.

Tutkimusaineistojen perusteella molemmat tutkimuksen kohteena olevat substantiivit muodostavat useimmin yhdyssanoja muiden lekseemien kanssa EU-aineistossa kuin SL-verrannaisaineistossa suhteessa aineistojen kokoon.

Toimenpide-yhdyssanoja on EU-aineistossa vähän tyyppejä, mutta paljon sanaesiintymiä, kun taas SL-aineistossa on vähän molempia. Yksinään sane *toimenpide* esiintyy EU-aineistossa huomattavasti SL-aineistoa taajemmin (tarkemmin alaluvussa 6.1).

Taulukko 48. Yhteenvedo: tutkittavia substantiiveja sisältävien yhdyssanojen lukumäärät aineistoissa

	EU tyyppi/sana- esiintymä	EU per 100000	SL tyyppi/sana- esiintymä	SL per 100000
<i>-toimenpide-</i>	96/613	8/50 (0,16)	102/490	3/13 (0,23)

Yhdyssanan yhteisesiintyminen muiden kielellisten elementtien kanssa samassa tekstiyhteydessä luo toisen asteen yhteisesiintymän. Ensimmäinen aste on yhdyssanan määräiteosan ja edusosan yhteisesiintymä. Yhteenvedotaulukossa 49 yhdyssanojen ryhmään on laskettu mukaan kaikki yhdyssanat, joiden edusosa on tutkittava substantiivi *toimenpide*.

Taulukko 49. Yhteenvedo: tutkittavien substantiivien esiintyminen yhdyssanan edusosana

lekseemi	EU lekseemi it- senäisenä	EU lekseemi yhdyssanan edusosana	SL lekseemi it- senäisenä	SL lekseemi yhdyssanan edusosana
<i>toimenpide</i>	2180	585	2763	445

Yhdyssanat ovat suhteessa aineiston kokoon selkeästi tavallisempia EU-aineistossa kuin SL-aineistossa. Voidaan arvella, että yhdyssanat ovat yleistymässä Suomen säädöskielen erikoiskiellisenä sananmuodostuskeinona.

Tulkitsen yhteisesiintymän kerrosteiseksi ja sisäkkäiseksi ilmiöksi. Yhteisesiintymät voivat kiteytyä yhdyssanoiksi, jotka kirjoitetaan useimmiten tai aina yhteen, jolloin ne ottavat ydinsanan roolin yhteisesiintyessään ympäristössään olevien sanojen kanssa. On mahdollista, että sama yhdyssana sen osat erikseen kirjoitettuna tuottaa toisenlaisia yhteisesiintymiä. Sisäkkäisyys tarkoittaa myös sitä, että yhteisesiintymä (vaikkapa samainen *suomen kieli*, jokin lauseke tai sanonta) voi puolestaan yhteisesiintyä erilaisten kielellisten elementtien sisällä, jolloin muodostuu syklisiä kerrostumia, esimerkiksi säädösteksteille tyypillisiä vakiolausekkeita tai vakiolauseita.

Tässä tutkimuksessa yhdyssanat on käsitelty erikseen ja niitä koskevat havainnot on tehty tulosten analyysin myötä, eivätkä yhdyssanat yhteisesiintyjinä siten ole mukana määrällisissä havainnoissa.

6.5.2 Läheisyys, vierekkäisyys, järjestys

Tutkimusverbien jakaumat EU-aineistossa hakusanan *toimenpide* positioihin eroavat toisistaan. Yhteisesiintymässä *tehdä* + *toimenpide* lähes kaikki verbin esiintymät ovat positioissa L4 (32 prosenttia), L2 (32 prosenttia) ja R5 (32 prosenttia), mikä johtuu aineistossa taajaan esiintyvistä vakiolausekkeista (esimerkki 24 sivulla 111). Havainnot semfinlex-lait-aineistosta eivät vastaa EU-aineiston tuloksia verbin *toteuttaa* osalta, joka jakautuu muita verbejä tasaisemmin positioihin ja esiintyy myös näiden ulkopuolella positioissa L6–L15 ja R6–R15.

Verbi *tehdä* esiintyy kyseisessä vakiolausekkeessa tutkimusikkunan sisällä kaikissa kolmessa edellä mainituissa positioissa EU-aineistossa. Aineiston semfinlex-lait tulokset noudattavat SL-aineiston jakaumaa verbin *tehdä* osalta, kun suurin osa yhteisesiintyjän havainnoista on positioissa L2 ja R1. Tutkimusikkunan ulkopuolelta on havaintoja positioista R6–R5 ja positioista L6–15, mikä on poikkeuksellista kaikissa tutkimusaineistoissa.

Säädöskielelle tyypilliset vakiolausekkeet ovat tavallisia yhteisesiintymässä *ryhtyä* + *toimenpide*. Se näkyy hyvin verbin jakaumassa eri positioihin tutkimusikkunassa ennen ja jälkeen hakusanan. Pääosa, lähes 45 prosentissa kaikista verbin *ryhtyä* esiintymistä SL-aineistossa yhteisesiintymän komponenttina sijaitsee positioissa L4, jolloin verbin ja hakusanan välissä on kolme sanetta. EU-aineistossa positiossa L4 ei ole lainkaan verbin *ryhtyä* esiintymiä, eikä tutkimusaineistossa siten myöskään ole SL-aineistolle tyypillisiä vakiolausekkeita, jollainen oprosenttian esimerkiksi ”ryhtyä lain täytäntöönpanon edellyttämiin toimenpiteisiin”. EU-aineistossa yhteisesiintymän *ryhtyä* + *toimenpide* verbin tyypilliset positiot ovat L1 ja L2 joissa molemmissa on yli 27 prosenttia esiintymistä, ja verbin ja hakusanan väliin mahtuu jälkimmäisessä tapauksessa esimerkiksi adjektiivi, useimmiten *tarvittava* tai *tarpeellinen*.

Verbi *toteuttaa* korvaa verbin *ryhtyä* EU-asetusten vakiolausekkeissa. Positiokohtainen tarkastelu ei kuitenkaan tarjoa riittävästi evidenssiä tukemaan ajatusta interferenssin vaikutuksesta yhteisesiintyjän valinnassa. Taaajimmin EU-aineistossa substantiivin *toimenpide* kanssa yhteisesiintyvä verbi *toteuttaa* esiintyy myös muita tutkimusverbejä useammin hakusanasta etäällä olevissa positioissa L5 ja R5, mutta kuitenkin huomattavasti vähemmän kuin positiossa L2, L1, R1 ja L3, joissa pääosa esiintymistä sijaitsee.

SL-aineistossa *suorittaa* ja *toteuttaa* puolestaan esiintyvät harvakseltaan, eikä esiintymiä missään positiossa ole kahta enempää. Verbien *säätää* ja *tehdä* jakaumat positioihin puolestaan ovat lähellä toisiaan molemmissa aineistoissa, kun taas verbien *ryhtyä* ja *soveltaa* jakaumat ovat selvästi toisistaan poikkeavia. EU-aineisto ja SL-aineisto osoittautuvat siis keskenään erilaisiksi, kun tarkastellaan tutkittujen yhteisesiintymien verbien sijoittumista positioihin. Positiokohtainen tarkastelu vah-

vistaa ennakkokäsitystä säädöskielen rakenteesta, jonka tyypillinen piirre on vakio-lausekkeiden runsaus. Erityisesti silloin, kun kaikki tai suurin osa verbikollokaateista on keskittynyt yhteen tai muutamaaan positioon, voidaan olettaa, että käytössä on vahva vakiolauseke.

Havainnot tukevat ajatusta siitä, että säädöskieli on luonteeltaan vakiolausekkeita hyödyntävää ja niihin pakkautuvaa, mutta EU-kieli ja Suomen säädöskieli poikkeavat toisistaan vakiolausekkeiden sisällön suhteen, mikä kuvastuu erilaisina yhteisesiintymärakenteina aineistoissa.

Yhteisesiintyjän etäisyys hakusanasta: vieruskollokaatit

Sinclair sanoi vuonna 2004 Wolfgang Teubertin haastattelussa, että 5 sanetta ennen noodia (hakusanaa) ja 4 sanetta sen jälkeen riittää tutkittavaksi alueeksi yhteisesiintymien havaitsemiseksi englannin kielessä. Hän toteaa kuitenkin, että sellaisten kielten kohdalla, joissa sanajärjestys on ”vapaa” voi olla tarpeen tarkastella pidempää saneiden jonoa. (Sinclair et al. (2004): xix). Sinclair tarkasteli kuitenkin yhteisesiintymän esiintymistä tutkimusikkunan sisällä, eikä juuri kiinnittänyt huomiota yhteisesiintyjien positiokohtaisiin esiintymiin. OSTI-raportin mukaan (Sinclair et al. (2004)) eniten yhteisesiintyjä löytyy noodin molemmin puolin, ja tutkimusikkunan/jatkumon reuna-alueilla niitä on selvästi vähemmän.

Tutkimusaineiston tutkimusikkuna on erikoiskielen ominaispiirteet huomioiden pieni, eikä tuloksista löydy käytetyllä menetelmällä lauserajoja ylittäviä tutkimuksen määritelmän mukaisia yhteisesiintymiä. Jos yhteisesiintymä määritellään assosiatiiviseksi saman aihealueen tai tilannekontekstin mahdollistamiksi leksikaalisiksi valinnoiksi, on säädöskieli hyvin todennäköisesti leksikaaliselta tiheydeltään tiivistä ja sisältää siten paljon ainesta, joka on mahdollista tulkita yhteisesiintymisen piiriin kuuluvaksi. Lainsäädännölliset käsitteet muodostaisivat siten assosiatiivisen verkoston, joka toimisi tekstien taustalla. Tämän ulottuvuuden kartoitus ei kuitenkaan ole tutkimukseni piirissä.

Sinclairin näkemys vieruskollokaattien vallitsevasta asemasta saa tukea tämän tutkimuksen tuloksista. Sekä tutkimusaineistossa että verrannaisaineistossa hakusanan lähellä olevat positiot saavat enemmän esiintymiä kuin muut positiot tutkimusikkunassa. Poikkeuksen muodostavat EU-aineistossa verbi *tehdä* ja SL-aineistossa verbi *ryhtyä*, joita esiintyessään *toimenpiteen* kanssa sijaitsee positioissa L4 positioiden L2 ja L1 lisäksi. *suorittaa* esiintyy molemmissa EU-aineistossa positiossa L1 ja SL-aineistossa positiossa R1 ja *toteuttaa* EU-aineistossa pääosin positiossa L2 ja SL-aineistossa positiossa R1. Hakusanan oikealla puolella positioissa R1-R5 esiintyvät verbikollokaatit ovat usein verbikantaisia substantiiveja, kuten *soveltaminen* tai *toimiminen*.

Raportoin tulokset positioittain, vaikka yhteisesiintyjän sijainti suhteessa hakusanaan ei ole sidoksen tai merkityksen kannalta ratkaiseva. Esiintymien laskenta positioittain motivoituu toisaalta vakiolausekkeiden löytymisellä niiden kautta ja toisaalta lukujen avulla voin verrata tutkimusaineistoa verrannaisaineistoon.

Lausekkeen tai virkkeen rajan ylittäviä yhteisesiintymiä

Ammattikieliä ja erikoiskieliä tutkittaessa voidaan olettaa, että hakusanan ja sen yhteisesiintyjän välissä voi olla paljonkin etäisyyttä. Tutkimieni aineistojen perusteella substantiivien ja verbien yhteisesiintymän tarkasteluun riittää kolme sanetta ennen hakusanaa ja kolme sen jälkeen. Reuna-alueilta positioissa L4/R4 ja L5/R5 löytyy verbiyhteisesiintyjien yksittäisiä esiintymiä, mutta suurin osa yhteisesiintyjistä on ryhmittynyt hakusanan ympärille sen kolmeen lähimpään positioon molemmin puolin. Tulos ei verbiyhteisesiintyjien tarkastelun perusteella osoita intuitiivista olettamusta yhteisesiintyjän rakenteellisesta kompleksisuudesta ja yhteisesiintymän hakusanan ja yhteisesiintyjän välillä olevan tekstistä perustelluksi.

6.5.3 Kahden yhteisesiintyjän rakenne

Jos samassa tutkimusikkunassa esiintyy useampi kuin yksi tutkimuksen kohteena olevista verbeistä *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä tai toteuttaa*, on niiden kanssa esiintyvänä toisena verbiyhteisesiintyjänä useimmiten verbi *toteuttaa*, kuten esimerkissä 51 alla.

On ryhtynyt toteuttamaan toimenpiteitä

Toteutettavien toimenpiteiden soveltamisen kesto

N:o 804/68 12 artiklassa säädetyt toimenpiteet, jotka toteutetaan maitotuotteiden ylijäämien

Esimerkissä 46 on kolme esimerkkiä rakenteesta $V + V + \textit{toimenpide}$ (EU-aineisto).

- (46) N:o 804/68 12 artiklassa säädetyt toimenpiteet, jotka toteutetaan maitotuotteiden ylijäämien

Edellä esitettyjen teknistä rakennetta valottavien havaintojen lisäksi tutkimusaineistosta nousee määrällisten havaintojen myötä esiin joitakin huomionarvoisia syntaktisia seikkoja, jotka liittyvät kieliopilliseen lukuun ja aktiivi/passiivi-rakenteisiin. Esitän niihin liittyvät havainnot alla.

6.5.4 Verbikantaiset substantiivit

Tässä tutkimuksessa lähdettiin oletuksesta, että kollokaation pääsana on substantiivi ja kollokaatti on verbi. Kuitenkin tutkimuksen kuluessa verbin kieliopillinen muoto alkoi näyttäytyä toissijaiselta kriteeriltä merkityksen ottaessa ensisijaisen kriteerin paikan. Myötäesiintyjä voi siten motivoitua myös syntaktisten ehtojen perusteella, muun muassa (Sinclair (1991), Sinclair et al. (2004)), jolloin liikutaan sanaston ja kieliopin raja-alueella, jossa ilmiöt näyttäytyvät limittäisinä tai päällekkäisinä. Laadullisessa analyysissä olen ottanut mukaan laskelmiin verbikantaiset substantiivit, koska ne tuottavat saman merkityksen kuin tutkimuksen kohteena olevat verbikollokaatit. Kun verbit on perusmuotoistettu, ovat kaikki verbimuodot mukana laskennassa, eikä kontekstuaalisen tiedon valossa näyttäisi olevan mitään syytä jättää verbikantaisia substantiiveja tutkimuksen ulkopuolelle. Siten mukaan on laskettu *toteuttaa toimenpide*, *toimenpiteen toteuttaminen*, *toimenpiteen toteutus* ja *toimenpiteen toteuttamisajankohta* ja *toteutustapa*.

Havainnot verbikantaisista substantiiveista on esitetty seuraavassa tutkimusverbeittäin. Poikkeuksena on edellä aluvuossa 6.1 *tehdä*-verbin kuvauksessa esitetty esimerkki 24 verbistä *tehdä*, jonka kanssa samassa yhteisesiintymässä V + *toimenpide* useimmiten esiintyvä verbi on *tehdä*. Verbikantaiset substantiivit muodostavat usein yhdyssanoja, kuten esimerkiksi verbin *soveltaa* kuvauksessa alla.

a) *ryhtyä*

Tutkimusaineistossa ja verrannaisaineistossa verbin *ryhtyä* verbikantaisia substantiiveja esiintyy vähän. EU-aineistossa on yhteensä 40 hakusanan *toimenpide* ja verbin *ryhtyä* yhteisesiintymää, joista kuudessa (15 prosentissa) esiintyy verbikantaisena substantiivina *ryhtyminen*. Verrannaisaineistosta SL löytyy samoin vain 41 esiintymää, joissa *ryhtyä* on verbikantaisen substantiivin muodossa (4,2 prosentissa yhdistelmistä *ryhtyä* + *toimenpide*). SL-aineistossa löydösten pieni prosentuaalinen osuus selittyy osittain vakiofraasien ja niiden modifikaatioiden suurella määrällä.

Alla on esimerkki 47 verbikantaisesta substantiivista *ryhtyminen* EU-aineistossa.

- (47) Valitukset saattavat helpottaa komission ryhtymistä toimenpiteisiin, kun ETY:n perustamissopimuksen 85 (EU)

Yhdyssanoja, joiden määrittäjäosa on verbikantainen substantiivijohdos *ryhtymis-* ei löytynyt lainkaan. *Ryhtymis-* saattaa muutenkin olla suhteellisen harvinainen yhdyssanan määriteosana. Internetistä löytyy googlen hakukoneella muutama ehdokas, esimerkiksi *ryhtymis- ja identiteettikysymys* ja *ryhtymisosa*.

b) soveltaa

Kuten jo aikaisemmin on huomattu, suuri osa verbin *soveltaa* esiintymistä osana tutkittua yhteisesiintymää EU-aineistossa on verbikantaisia substantiiveja (esimerkki 47 yllä), esiintymiä on 198 havainnon joukossa yhteensä 112 kappaletta. Pelkääntään verbikantaisia substantiiveja *soveltaminen* esiintyy V+N -rakenteessa *suorittaa + toimenpide* peräti 98 kertaa. EU-aineistossa on myös suhteellisen paljon havaintoja, joissa verbi sijoittuu osaksi yhdyssanaa sen määräiteosaksi. Esimerkiksi sanaliitto *toimenpiteiden soveltamisala* avautuu yhteisesiintymäksi *soveltaa + toimenpide + ala*, joka rakentuu yhdyssanan avulla muotoon *toimenpiteen soveltamisala*. Katso esimerkki 48. Se voidaan tällöin myös lukea verbikantaisiin substantiiveihin tutkittaessa yhdyssanoja kollokaatioiden edustajina. Esiintymiä on yhteensä 14 kappaletta ja ne sijoittuvat pääosin positioon R1.

Esimerkissä 48 kuvataan yhteisesiintyjä *soveltaa* verbikantaisena substantiivina positioissa R1 ja R5. (EU)

- (48) tosiasioiden perusteella on perusteltua laajentaa toimenpiteiden
6 soveltamisalaa, neuvosto päättää siitä (EU)
Yhteisön oikeudessa säädetään niiden hallinnollisten toimenpiteiden
ja seuraamusten luonteesta ja soveltamisalasta (EU)

SL-aineiston viiden yhteisesiintymän *soveltaa + toimenpide* joukossa ei ole yhtään verbikantaista substantiivia.

c) suorittaa

V+N -rakenteissa *suorittaa + toimenpide* on yleisesti ottaen EU-aineistossa erittäin harvinainen osana tutkittua yhteisesiintymää, ja 23 esiintymän joukossa on vain yksi verbikantainen substantiivi *suorittaminen*. Verrannaisaineistossa SL puolestaan rakenne on huomattavasti tavallisempi ja aineistosta löytyykin verbikantaisia substantiiveja *suorittaminen* yhteensä 73 kappaletta, joiden lisäksi löytyy 5 kappaletta *suorittamis-* ja *suoritus-*alkuisia yhdyssanoja, esimerkkeinä *suorittamispaikka*, *suorituspaikka* ja *suoritusajankohta*.

d) säätää

Verbikantaisia yhteisesiintyjjiä, kuten *säätäminen*, *säätäjä*, ei ole EU-aineistossa eikä SL-aineistossa itsenäisinä leksemeinä eikä myöskään yhdyssanan määräiteosina osana tutkittuja yhteisesiintymiä. Niiden tyypilliset positiot olisivat alueella R1–R5, mutta näissä positioissa ei ole lainkaan verbin *säätää* esiintymiä.

e) *tehdä*

Verbikantaisista substantiiveista verbin *tehdä* muotoa *tekeminen* itsenäisenä lekseminä ei löydy yhtään esiintymää osana tutkittua yhteisesiintymä varsinaisessa tutkimusaineistosta (EU) eikä myöskään ensimmäisessä verrannaisaineistosta (SL). EU-aineistossa ei esiinny verbin *tehdä* verbikantaisia substantiiveja yhdyssanan määriteosina, mutta SL-aineistossa niitä on yksi kappale, joka on esitetty esimerkissä 50.

Esimerkissä 49 verbikantainen substantiivi on yhdyssanan määriteosa. (SL)

- (49) Annetun lain 9 §:n mukaisten toimenpiteiden teettämiskustannusten takaisinperinnän valtiolle suorittaa asianomainen (SL) || Rivi 590247

f) *toteuttaa*

EU-aineistosta osana tutkittavia yhteisesiintymiä löytyy 11 yhdyssanaa, joiden määriteosat ovat verbikantaisia substantiiveja ja joiden kanta löytyy tutkittavien verbien joukosta. Seitsemän niistä on *toteuttamis*-alkuisia, ja neljä *toteutus*-alkuisia. Esitän jälkimmäisestä esimerkin 50. Aineiston yhteisesiintyjien joukossa on myös yksi itsenäisesti esiintyvä verbikantainen substantiivi *toteutus* ja 56 verbikantaista substantiivia tyyppiä *toteuttaminen*.

- (50) käytetyksi, jos hankkeen tai toimenpiteen toteutus edistyy suunnitellulla tavalla ja (EU)

Yhteenveto: verbikantaiset substantiivit

Varsinainen tutkimusaineisto (EU) ja ensimmäinen verrannaisaineisto (SL) ovat verbikantaisten substantiivien osalta samanlaisia verbien *säätää* ja *tehdä* suhteen: kumpikaan verbi ei esiinny verbikantaisena substantiivina tutkimuskohteen *toimenpide* kanssa tutkimusikkunassa. Sen sijaan verbit *soveltaa* ja *suorittaa* käyttäytyvät aineistoissa vastakkaisilla tavoilla. EU-aineistossa suurin osa yhteisesiintymän *soveltaa* + *toimenpide* yhteisesiintyjistä on verbikantaisia substantiiveja, kun taas SL-aineistossa tämä on tavallista yhteisesiintymässä *suorittaa* + *toimenpide*. EU-aineistossa tavallinen yhteisesiintyjä *toteuttaa* on usein verbikantainen substantiivi. Tutkimusverbi *ryhtyä* puolestaan esiintyy vähän verbikantaisena substantiivina molemmista aineistoista, sen sijaan se osallistuu usein vakiolausekkeiden muodostamiseen.

Verbikantaiset substantiivit yhteisesiintyjänä

Verbikantaiset substantiivit ovat yliedustettuja tutkimusaineistossa (EU) verrattuna ensimmäiseen verrannaisaineistoon (SL) tutkittavien substantiivien yhteisesiintymien joukossa. Syytä on vaikea selvittää tietämättä kyseisen tekstin lähdetekstin rakennetta, joten interferenssin vaikutusta ei voi tässä tutkimuksessa sulkea pois, muttei myöskään vahvistaa. Substantiivin suosiminen kokonaisen sivulauseen tai kokonaisen lausekkeen sijaan tuottaa kuitenkin tiiviimpää kieltä, joka on todennäköisesti ollut käännösvaiheessa tavoitteena.

Olen koonnut alla olevaan taulukkoon 50 verbikantaisten substantiivien lukumäärät tutkituissa yhteisesiintymissä. Taulukossa lekseemien kokonaismäärät aineistoissa on ilmoitettu otsikkorivillä. Jakoviiva erottaa itsenäisenä lekseeminä ja yhdyssanan osana esiintyneiden verbikantaisten substantiivien lukumäärät. Verbikantaisten substantiivien lukumäärät sisältävät yhdyssanojen määriteosana esiintyneet verbikantaiset substantiivit. Verbikantaisten substantiivien jakaumat termin *toimenpide* yhteisesiintymissä noudattavat pääpiirteissään verbien jakaumia, jolloin verrannaisaineistossa tavalliset *ryhtyä* ja *suorittaa* ovat tavallisia myös verbikantaisina substantiiveina *ryhtyminen* ja *suorittaminen*. EU-aineistossa tosin taajimmin esiintyvä verbiyhteisesiintyjä *toteuttaa* häviää tässä suhteessa poikkeuksellisesti verbikollokatile *soveltaa*, joka esiintyy verbikantaisena substantiivina noin kaksi kertaa lekseemiä *toimenpide* useammin.

Verbikantaisten substantiivien muoto vaihtelee, mutta tavallisimpia ovat *-minen*-loppuiset substantiivit *ryhtyminen*, *soveltaminen*, *suorittaminen*, *säätäminen*, *tekeminen* ja *toteuttaminen* sekä variaatiot *suoritus* ja *toteutus*. Myös tekijää ilmaisevia muotoja, kuten *suorittaja* ja *toteuttaja*, esiintyy verbikantaisten substantiivien joukossa jonkin verran.

6.5.5 Kieliopillinen luku

Hakusana *toimenpide* on yhteisesiintymien osana EU-aineistossa lähes aina monikossa. Toimenpide on EU-aineistossa yksikkömuotoinen 33 tapauksessa 2 180 esiintymästä. Yhteisesiintyjien joukossa poikkeuksellinen on verbi *suorittaa*, joka esiintyy yli 21-prosenttisesti yksikkömuotoisen hakusanan kanssa EU-aineistossa. Sama tendenssi näkyy verrannaisaineistossa SL, jossa ainoa selkeä poikkeus on samoin verbi *suorittaa*, joka esiintyy lähes 70-prosenttisesti yksikkömuotoisen hakusanan kanssa. Taulukkoon 51 on koottu yksikkömuotoisten hakusanojen *toimenpide* havaintojen lukumäärät ja prosentuaaliset osuudet yhteisesiintymissä tutkimusverbien kanssa.

Ylivoimainen enemmistö kaikista tutkittavista substantiiveista on monikossa. *toimenpide* on lähes aina monikossa tutkimusverbien kanssa esiintyessään, lukuun ot-

Taulukko 50. Verbikantaisten substantiivien lukumäärät yhteisesiintymissä.

<i>toimenpide</i>	EU	SL
itsenäisenä	2180	2763
yhdyssanan osana	585	445
verbi	EU	SL
<i>ryhtyä</i>	6/6	41/7
<i>soveltaa</i>	126/33	0/3
<i>suorittaa</i>	1/1	78/8
<i>säätää</i>	0	0/1
<i>tehdä</i>	0	1
<i>toteuttaa</i>	68/11	29/10
yhteensä	201/51	120/30

tamatta verbiä *suorittaa*, jonka yhteydessä *toimenpide* on yksikössä noin 20 prosentissa esiintymistä. Verrannaisaineistossa (SL) *toimenpide* esiintyy 70-prosenttisesti yksikössä verbin *suorittaa* yhteydessä, muutoin aina monikossa.

Taulukko 51. *toimenpide* yksikössä. (EU ja SL)

verbiyhteisesiintyjä	EU kpl	EU %	SL kpl	SL %
<i>ryhtyä</i>	1	2,5	100	1,0
<i>soveltaa</i>	2	1,0	0	0,0
<i>suorittaa</i>	5	21,7	207	69,0
<i>säätää</i>	0	0,0	5	29,4
<i>tehdä</i>	1	2,5	45	6,6
<i>toteuttaa</i>	3	5,3	31	3,4

Verbiyhteisesiintyjät *soveltaa* ja *toteuttaa* esiintyvät EU-aineistossa osana yhdysanaa, jossa verbikantainen substantiivi muotoa *soveltamis-* tai *toteuttamis-* määriteosassa.

Prosentuaaliset erot molemmissa aineistoissa viittaavat siihen, että *toimenpide* yksikössä tuottaisi sellaisen kieliopillisen kontekstin, jossa verbiksi valikoituisi *suorittaa*. Termin luku puolestaan on riippuvainen tekstinulkoisesta todellisuudesta, tilannekontekstista.

Kolmannessa verrannaisaineistossa (LaDi) määrällisiä havaintoja voi verrata suoraan varsinaisen tutkimusaineiston ja ensimmäisen verrannaisaineiston tuloksiin vain asetusten osalta, joita kuitenkin on aineistossa vain neljä tekstiä. Kun ottaa huomioon, että direktiivit käyvät läpi sekä tekstin että sisällön tarkistus- ja muokkausvaiheen, voisi olettaa, että direktiivit olisivat LaDi-aineistossa enemmänkin lakien kaltaisia. Tulosten valossa näin ei kuitenkaan ole, direktiivit ovat lähempänä EU-asetuksia kuin kansallisia lakeja määrällisten tulosten perusteella.

6.5.6 Epäsuorat objektit

Yhteisesiintymän komponentti, joka laskennassa tavallisesti saisi yhteisesiintyjän aseman esiintymistääjuutensa perusteella, osoittautuukin epäsuoraksi objektiksi, joka ei merkityksensä puolesta liitykään tutkimusverbiin, eikä ole hyväksyttävä esiintymä. Näitä tavanomaisia yhdistelmiä tutkimusaineistossa ovat esimerkiksi tyyppin *soveltaa x toimenpiteisiin* vakiorakenteet, kuten esimerkissä 51. Siinä sovellettava asia on lain kohta, ei toimenpide. Esimerkissä 52 esitetään havainto, joka on hyväksytty analysoitavaksi. Vastaavia rakenteita, joissa verbiyhteisesiintyjä on tutkittujen verbien joukossa, löytyy kaikkien tutkittujen substantiivien tekstiyhteydestä.

(51) 2 Edellä 1 kohtaa **sovelletaan toimenpiteisiin**, jotka a) ovat (EU)

(52) N:o 804/68 12 artiklassa säädetyt toimenpiteet, jotka toteutetaan maitotuotteiden ylijäämien (EU)

Tutkimuksen valossa yhteisesiintymä voidaan tulkita sanaluokkien rajoja ylittäväksi ilmiöksi. Näkemys vähentää sanaluokan ja nostaa dependenssijäsenyyksen merkitystä hakukriteerinä etsittäessä kollokaatioita aineistosta.

Syntaktisten konstruktioiden välisiä yhteisesiintymiä nousee esiin jo 11 saneen tutkimusikkunassa. Seuraavassa esimerkissä yhdyssanan *toteutus selvitys* määriteosa viittaa konstruktion **toimenpiteen toteutus**, ja sen voidaan näin laskea yhteisesiintymäksi. Esimerkissä 53 on kahden syntaktisen konstruktion yhteisesiintymä SL-aineistosta.

(53) Muiden toimenpiteiden rahoituksen tulee perustua toimenpiteen toteutuksen jälkeen annettavaan **toteutus selvitykseen** (SL)

Luku 7

Merkitykseen ja assosiaatioon liittyvät havainnot: *toimenpide*

Tässä luvussa esitän ne merkitykseen ja assosiaatioon liittyvät havainnot, jotka ovat nousseet esiin tutkimuskohteen *toimenpide* ominaispiirteiden määrällisen tarkastelun yhteydessä (ks. luku 6). Tavoitteena on tarkentaa niiden perusteella kuvaa kollokaatiosta ilmiönä hahmottelemalla sen merkitykseen ja assosiaatioon liittyviä ominaispiirteitä siinä määrin, kun tutkimuksessni on osoittautunut mahdolliseksi. Yhteisesiintymä on potentiaalinen vakiintuma, mikäli vakiintumalle asetetut kriteerit täyttyvät. Yhteisesiintymän ja vakiintuman käsitteet ja ominaispiirteet on kuvattu luvussa 2.

Distributionaalisessa tarkastelussa lähtökohtana on, että kollokaation elementit, hakusana ja yhteisesiintyjä, esiintyvät usein yhdessä. Fraseologisen lähtökohdan kriteerinä yhteisesiintymän luokittelussa vakiintumaksi on kollektiivinen merkityksen muodostuminen. Kuten merkityspiirteitä kartoittavassa alaluvussa 2.2.3 totesin, näen merkitykseen liittyvät ominaispiirteet ensisijaisesti kollokaation muodostumisen motiivina. Tutkimukseni aineisto ei tarjoa mahdollisuutta merkityksen tai assosiaation laaja-alaiseen tutkimukseen, joka edellyttäisi omalla aineistollani enemmän laadullista tutkimustyötä kuin mitä tutkimukseni puitteissa on ollut mahdollista. Tulosten kvantitatiivisen käsittelyn tuloksista on tutkimuksen kuluessa kerääntynyt havaintoja kollokaation merkitykseen liittyvistä ominaispiirteistä. Näin kysymyksenasettelu on pysynyt tutkimuksessa mukana. Liittymäkohtia kvantitatiivisen ja merkitykseen sitoutuvan tarkastelun välillä ovat muun muassa vastavuoroisuuden käsite, kompositionaalisuus ja merkityksen muodostuminen yhteisesiintymisen kautta. Leksikaalisen virittämisen teoria mahdollisesti selittää osaltaan uusien yhteisesiintymien vakiintumisen kieliyhteisössä ajan mittaan.

Esitän merkitykseen liittyvät havainnot aloittaen tutkimuskohteen *toimenpide*

Haettiin ”toimenpide”Kielitoimiston sanakirjasta [Ohjeet] [Tietoa sanakirjasta] [Anna palautetta sanakirjasta]
 toimenpide48*F jnk määrätehtävän suoritus, teko; vars. monikossa us. paremmin: toimi. Uuden johtajan ensimmäinen toimenpide oli hankkia sihteeri.
 Turvallisuustoimenpiteet paremmin: turvatoimet. Ryhtyä toimenpiteisiin paremmin: toimiin.
 Lääk. Kirurginen toimenpide. Potilaalle tehty hoitotoimenpiteet.

Kuva 7.1: *toimenpide* Kielitoimiston verkkosanakirjassa, kuvakaappaus verkkosivulta www.kotus.fi

tarkastelusta EU-tekstien suomentamisen kontekstissa selvittämällä aluksi sen käännösvastineita Euroopan unionin termipankista IATE⁶⁰, joka on Euroopan unionin entisen termipankin Eurodicautomin seuraaja. Pysin hahmottamaan termin käyttöä tutkimalla käännösvastineita takaisinkääntämisen menetelmällä. Haen tietokannasta ensin suomenkielisen termin englanninkieliset vastineet ja sen jälkeen näiden englanninkielisten vastineiden IATEen kirjattuja suomenkielisiä vastineita. Termi-tietokannasta selviää myös, minkä lainalan termi kyseinen vastine EU:n tekstien luokituksen valossa on. Tuloksena on käsitys termin *toimenpide* käytöstä eri EU:n lainsäädännön alueilla 2010-luvulla.

Tarkastelen myös lyhyesti semanttiseen prosodiaan ja preferenssiin liittyviä havaintoja ja pohdin teoriaa merkittävistä ja tavallisista yhteisesiintymistä tutkimusaineiston näkökulmasta. Kartoitan yhteisesiintymän tekstiympäristöstä viitteitä tyyppillisistä aihealueista LaDi-aineistossa osana termin *toimenpide* merkityksen tarkastelua säädöskielessä. Muista aineistoista ei vastaavaa tarkastelua ole ollut mahdollista tehdä aineistojen laajuuden vuoksi.

Kotuksen verkkosanakirjassa (kuva 7.1) hakusana *toimenpide* asettuu esimerkkien perusteella lääketieteen ja mahdollisesti yritysmaailman ammattikieleen, jolloin sanakirjan mukaan verbiyhteisesiintyjä ensiksi mainitussa kontekstissa olisi *ryhtyä* ja jälkimmäisessä *tehdä*. Havainnot verbistä *tehdä* yhteisesiintymässä hakusanan *toimenpide* kanssa lääketieteen erikoiskielessä löytyvät luvusta 6. Erityisesti monikkomuotoisena *toimenpide* on Kotuksen sanakirjan mukaan parempi korvata (lähi)synonyymillä *toimi*, mikä tarkoittaisi tutkimusaineistoissa kaikkien verbien yhteisesiintymiä, lukuunottamatta verbiä *suorittaa* EU-aineistossa ja SL-aineistossa ja verbiä *säätää* SL-aineistossa.

⁶⁰iate.europa.eu

7.1 Käännösvastineita EU:n IATE-termipankissa

Yleisesti ottaen voidaan sanoa, että *toimenpide* on EU-aineistossa usein käytetty termi (ks. alaluku 6.1). IATE-termipankissa vakiintuneita englanninkielisiä vastineita löytyy oikeusaloilta⁶¹ oikeus, finanssi, hallinto, kauppa ja ympäristö.

Katsauksen tavoitteena on hahmottaa, mitä käännösvastineiden vaihtoehtoja IATE tarjoaa englannista suomentavalle kääntäjälle, kun kyseessä on lähdetekstin termi, jonka yhtenä vastineena on tutkimuskohde *toimenpide*. Näin kääntäjän tekemä valinta selittyisi termipankkiin kirjattujen vastineiden kontekstin kautta. Tutkimuksessa ei ole pystytty selvittämään, mitä vastineita Eurodicautom on 1990-luvun puolella tarjonnut kääntäjän avuksi. Tietokantaan on linkitetty tarjotun vastineen tilannekontekstietuna termin lähdeteksti, jonka metatiedoista on mahdollista selvittää muun muassa tekstin alalaji ja vuosiluku. Vaikka asetuskohdaista tietoa tutkimusaineiston asetusten lähdekielestä ei ole saatavissa, tarjoaa 2010-luvulla tehty katsaus takaisinkäännöksiin kieliparissa suomi-englanti-suomi kuitenkin käsityksen termin *toimenpide* vastineiden kirjosta ja pääsyn lähemmäksi sen merkitystä EU-kontekstissa.

Esitän katsauksen ensimmäisen osion tulokset taulukossa 52. Hakutuloksissa tavallisin vastine on *measure*. Se on myös ainoa termille *toimenpide* annettu vastine Hallinnon kehittämiskeskuksen vuonna 1996 julkaisemassa sanastossa *Termit tutuiksi*⁶². Se on siten todennäköisin englanninkielinen termi, jonka vastineeksi kääntäjä on valinnut termin *toimenpide*. Sen lisäksi löytyy kuusi muuta vastinetta, jotka ovat *act*, *action* (teko, toiminta, toimenpide), *manoeuvre*, *measure*, *procedure*, *transaction* (finanssi) ja *undertaking*. Havainto tukee termin *toimenpide* produktiivisuutta käännösvastineena.

Tarkastelen seuraavaksi katsauksen tuloksia verrattuna varsinaisesta tutkimusaineistoni (EU) tuloksiin. Olen koonnut havainnot taulukkoon 53. Lihavoidut käännösvastineet esiintyvät varsinaisessa tutkimusaineistossani (EU) kyseisen verbin kontekstissa: esimerkiksi *määräys* on eräs verbin *soveltaa* kontekstissa esiintyvistä sanoista. Taulukkoon on myös kirjattu intuitiivinen ennusteeni siitä, mikä tutkimusverbeistä parhaiten sopisi englanninkielisen termin yhteisesiintyjäksi kyseisellä erikoisalalla.

Tutkimusverbit ovat molemmissa aineistoissa taajaan esiintyviä verbejä, erityisesti *soveltaa* tutkimusaineistossa (EU) ja *säätää* verrannaisaineistossa (SL). Vähemmän taajaan esiintyvien verbien tekstikontekstista ei ole havaittu toistuvia aihealueita. Yhteenvetona havainnoista listaan verbien aihealueita alla.

⁶¹luokituksena IATE:n oma luokitusjärjestelmä

⁶²Hallinnon kehittämiskeskus. 1996. Termit tutuiksi. Eurooliittisia sanoja ja sanontoja. Suomi-englanti-suomi. Helsinki: Edita Oy

Taulukko 52. Hakusanan *toimenpide* englanninkieliset käännösvastineet IATE-termipankissa ja kyseisten käännösvastineiden takaisinkäännösten vaihtoehdot suomeen päin käännettäessä.

	act	action	manoeuver manööveri	measure	procedure proseduuri	transaction transaktio	undertaking
asiakirja	x						
järjestely				x			
kanne		x					
kauppa						x	
keino				x			
laki	x						
liike			x				
liiketoimi, transaktio						x	
liikkumavara			x				
menetelmä					x		
menettely				x	x		
menettelytapa					x		
määräys	x						
oikeustoimi						x	
ohjelma		x		x			
seuraamus				x			
sitoumus							x
sitoutuminen							x
sopimus	x					x	
sovintoratkaisu						x	
suoritus					x		
säädös	x						x
tapahtuma						x	
teko	x	x					
toimenpide	x	x	x	x	x	x	x
toimi		x		x		x	
toiminta	x	x				x	x
toimintamuoto					x		
toiminto		x					
tosite						x	
yrittäjä							x

- EU-aineistossa verbi *soveltaa* liittyy lainsäädäntöön. Esimerkkejä kontekstista: *lainsäädäntö* (tavallisin), *laki*, *asetus*, *säännös*, *pykälä* ja *sopimus*.
- EU-aineistossa verbi *suorittaa* liittyy talouselämään. Esimerkkejä kontekstista: *maksu*, *vähennys*, *etuus*, *korvaus*, *osto* ja *meno*. Se liittyy myös työntekoon. Esimerkkejä kontekstista: *työ*, *palvelu*, *kuljetus* ja *valvonta*.
- EU-aineistossa verbi *toteuttaa* on niin laaja-alainen, että se ei profiloidu millekään osa-alueelle erityisen tyypilliseksi.
- SL-aineistossa verbi *ryhtyä* liittyy työskentelyyn ja rakentamiseen.
- SL-aineistossa verbi *soveltaa* liittyy lainsäädäntöön. Esimerkkejä kontekstista:

lainsäädäntö (tavallisin), *laki*, *asetus*, *säännös*, *pykälä* ja *sopimus*

- SL-aineistossa verbi *suorittaa* liittyy talouselämään. Esimerkkejä kontekstista: *maksu*, *vähennys*, *etuus*, *korvaus*, *osto* ja *meno*. Se liittyy myös työntekoon. Esimerkkejä kontekstista: *työ*, *palvelu*, *kuljetus* ja *valvonta*. Se liittyy myös vaaleihin. Esimerkkejä kontekstista: *vaali* ja *äänestys*
- SL-aineistossa verbi *säätää* liittyy prosesseihin. Esimerkkejä kontekstista: *poikkeus* ja *tuotantotuki*
- SL-aineistossa verbi *tehdä* liittyy pienempiin prosesseihin. Esimerkkejä kontekstista: *päätös*, *sopimus*, *sitoumus*, *ilmoitus*, *esitys* ja *valitus*
- SL-aineistossa verbi *toteuttaa* liittyy laajempiin prosesseihin. Esimerkkejä kontekstista: *hanke*, *ohjelma*, *valvonta*, *tarkoitus*, *pyrkimys* ja *päämäärä*

Taulukko 53. EU:n IATE-termipankin hakusanalle *toimenpide* tarjoamia käänkösvastineita sekä oletuksia niiden mahdollisista verbikollokaateista tutkimustulosten perusteella, poislukien *toteuttaa*, joka EU-aineistossa esiintyy kaikissa konteksteissa.

IATE: toimenpide	vastineita	hypoteesi, kollokaatteja olisivat
act	asiakirja, laki , määräys , sopimus, säädös	soveltaa (EU+SL), säätää
action	kanne, ohjelma, teko, toimenpide, toimi, toiminto	toteuttaa (SL)
manoeuvre	liike, liikkumavara, manööveri	tehdä (SL)
measure	järjestely, keino, menettely, ohjelma, seuraamus, toimenpide, toimi	toteuttaa (SL)
procedure	menetelmä, menettely, menettelytapa, suoritus, toimintamuoto	tehdä (SL)
transaction	kauppa , liiketoimi, transaktio, oikeustoimi, sopimus , sovintoratkaisu, taphtuma , toimenpide , tosite, toimi	suorittaa (EU+SL), tehdä, toteuttaa (EU)
undertaking	sitoumus, sitoutuminen, säädös, toimenpide, yritys	tehdä (SL)

Tutkimuskohde *toimenpide* on potentiaalinen vastine monenlaisille ilmauksille

Tutkimuskohteen *toimenpide* takaisinkäänköksiä tarkasteltaessa toinen kieli on englanti. Takaisinkäänkösten tarkastelusta selviää, että *toimenpide* hoitaa laajaa merkityskenttää ja tarjoutuu vastineeksi useille eri alojen termeille. Tutkimuksen perusteella voidaan perustellusti arvella, että *toimenpiteen* merkitys myös Suomen säädöskielessä tulee EU:n myötä laajentumaan. IATE:n rooli tutkimukseni käänkösvastineiksi

ehdolla olevien termien tarkastelussa on havaintoja vain osittain selittävä ja spekulatiivinen. Sen sisältämät vastineet ovat kertyneet ajan mittaan tehtyjen käännösten perusteella, joten kyseessä on kehitys, jossa alun perin tehdyt ratkaisut vaikuttavat myöhemmin tarjolla olevien vaihtoehtojen valikoimaan. Samalla tavoin EU-tekstien valinnat tulevat vaikuttamaan Suomen säädöskieleen.

Johtopäätöksenä esitän hypoteesin, jonka mukaan termin *toimenpide* laaja merkityskenttä selittyy ainakin osin sen käännösvastineiden kirjon laajuuden kautta, mikä osaltaan selittää verbin *toteuttaa* yliedustuksen varsinaisessa tutkimusaineistossa (EU). *Toteuttaa* näyttäisi asettuvan neutraaliksi verbiksi muiden tutkimuksessani käsittelemieni verbien joukossa ja soveltuvan siten monenlaiseen kontekstiin.

7.2 Yhdyssana yhteisesiintymänä

Kollegiaalisen merkityksen muodostumisen kriteerinä on kompositionaalisuuden väheneminen. Yhteisesiintyjä muuttuu vakiintumaksi, kun kollokaation merkitys ei enää ole täysin johdettavissa osiensa merkityksestä. Tarkastelen tässä alaluvussa tilannetta, jossa yhteisesiintyjä on jo kiteytynyt yhdyssanaksi, jolloin johtopäätökseni on, että kyseessä on kollokaatio ja mahdollisesti myös vakiintuma.

Yhdyssana on kiteytynyt niin pitkälle, että se kirjoitetaan suomenkielessä yhteen ja se yhteisesiintyy edelleen muiden kielellisten elementtien kanssa. Yhdyssanan muodostus nähdään tässä tutkimuksessa yhteisesiintyjän ja perusosan yhteisen merkityksen tuottamisen mekanismina. Kun yhteisesiintyminen näin toteutuu ja syntyy kollektiivinen merkitys, on tuloksena merkittävä vakiintuma, jonka yhteen kirjoitetulle muodolle ei todennäköisesti ole vaihtoehtoista erikseen kirjoitettua versiota. Jos yhdyssana pysyy kompositionaalisenä tai on jatkumossa matkalla kollegiaalisen merkityksen suuntaan, näen yhteen kirjoittamisen seurauksena vaihtelusta, joka on riippuvaista kielen puhujan omaksumasta käsityksestä yhteisesiintymän vakiintumisen asteesta. Yhdyssanan luominen kuvastaa myös uuden termin tarvetta ja on käännöstoimen tilannekontekstissa EU-aineiston valossa usein käytetty strategia. Yhteisesiintymisen tuloksena tapahtuva yhdyssanojen muodostuminen tuo suomalaiseen säädöskieleen uutta ainesta. Osa yhdyssanoista on ei-kompositionaalisia ja sumeita (sumea kuvaa tässä läpinäkyvän vastakohtaa), mikä voi johtua siitä, että yleisesti ottaen EU:n toimintamallit ja sen säätelyn kohteet yhteiskunnassa ovat kansalaisille tuntemattomia. EU-asiantuntijalle termi *interventiovoi* (ks. tarkemmin alla) saattaa olla hyvin läpinäkyvä ja selkeä termi.

Yhteisesiintyjä rajaa yhdyssanan edusosan merkityskenttää kattamaan kontekstinsa sallimia tulkintoja. Kuten Mäkisalo on suomenkielen yhdyssanoja koskevassa tutkimuksessaan osoittanut, millään yhdyssanojen tunnistamiseen vaikuttavis-

ta tekijöistä (semanttinen läpinäkyvyys, muodollinen vastaavuus ja osien järjestys) ei ole yksinään ratkaisevaa merkitystä yhdyssanan tunnistamisessa, vaan tunnistamisen prosessiin vaikuttaa useita samanaikaisia tekijöitä (Mäkisalo (2000)). Koska EU-kontekstissa on usein kyse suomalaisille uusista käsitteistä, jolloin ne ovat suhteellisen tuntemattomia, kunnes termi on integroitunut Suomen säädöskieleen. Semanttinen läpinäkyvyys ei erikoiskielessä välttämättä nouse ratkaisevaksi tekijäksi, kun ammattikieltä käyttää rajattu joukko alan asiantuntijoita. Useamman kuin kaksiosaisen yhdyssanan viimeinen osa ei välttämättä myöskään aina ole sen edusosa. Siten ei ole perusteita yksiselitteisesti rajata yhdyssanan edusosa kantamaan pääasiallista merkitystä. Toisin sanoen EU-aineiston yhdyssanojen jälkiosa ei aina välttämättä ole semanttinen edusosa ja alkuosa tai alkuosat semanttisia määriteosia (ks. johtopäätökset *interventio*-alkuisten yhdyssanojen analyysistä alla).

Yhteisesiintyjä tarkentaa tai muuttaa pääsanan merkitystä

Tässä osiossa tarkastelen yhdyssanan muodostumista yhteisesiintyjän *interventio* avulla. Analyysi tukee hypoteesia uuden aineksen integroitumisesta osaksi suomalaista säädöskieltä. Lekseemi *interventio* rajaa *toimenpiteen* merkityskentän kattamaan oman kontekstinsa sallimia tulkintoja. *Intervention* merkitystä on tosin kyseisen tekstin asiasisältöä tarkemmin tutkimatta vaikea määritellä ja arvelen, että se ei ole kansalaiselle helposti avautuva käsite. Kotuksen verkkosanakirjan mukaan *interventio* tarkoittaa **asioihin puuttumista**, mutta kun lekseemiä *interventio* tutkitaan tarkemmin sen lähikontekstissa EU-asetuksissa, sen kanssa kiinteän yhteisesiintymän toisin sanoen yhdyssanan muodostavat lekseemit tarjoavat merkityksestä monisäikeisemmän kuvan. Yhdyssanan perusosien tarkoitteet liittyvät jonkin tarvikkeen (*tuote, väline*) kauppaan ja kustannuksiin (*osto, kustannus, meno, hinta*) ja varastointiin (*varasto, makasiini, varastointi*). Yhdyssana *interventiohint*a saa myös määreeksen kahdessa tapauksessa tarkenteen *perus-* jolloin yhteisesiintymän muodostus noudattaa seuraavaa kollokaatioketjua:

hint + interventio/interventio + hinta = interventiohint
 interventiohint + perus = perusinterventiohint

Ketju ei aineiston perusteella voi muodostua osioista *perusinterventio* + *hint*a, koska aineistosta ei löydy yhtään itsenäistä lekseemiä *perusinterventio*. Lekseemi *interventio* voi määrittää myös ainesanaa *voi*, ja tuloksena olevan yhteisesiintymän *interventiovoi* kanssa yhteisesiintyy muun muassa substantiivi *erityismyynti*, jonka 8 esiintymästä aineistossa 3 on yhteisesiintymiä *interventiovo*in kanssa.

Säädösteksteissä verbillä *ryhtyä* yhteisesiintymän *ryhtyä* + *toimenpide* komponenttina ei ole sen perusmerkitykseen liittyvää toiminnan aloittamisen rajoitusta.

Haettiin ”interventio”Kielitoimiston sanakirjasta [Ohjeet] [Tietoa sanakirjasta] [Anna palautetta sanakirjasta]

interventio3

1. väliintulo, asioihin puuttuminen. Keskuspankin interventio valuutan devalvoitumisen hillitsemiseksi.
2. pol., oik. puuttuminen toisen valtion asioihin (pakkoa t. sen uhkaa käyttäen). Sotilaallinen interventio.
3. vars. lääk. toimenpide jolla pyritään vaikuttamaan yksilön t. jnk ryhmän käyttäytymiseen t. terveydentilaan.

Kuva 7.2: Interventio Kielitoimiston verkkosanakirjassa.

Mielestäni ei ole perusteltua tulkita verbin *ryhtyä* käyttämistä lainsäätäjän tarkoituksena säätää lakeja, joissa määritellään toiminnan aloittaminen riittäväksi lain kirjaimen täyttämiseksi. Yleiskielen merkitykseen liittyvä tulkinta voi selittää verbin käytön vähenemisen ja tilan antamisen verbille *toteuttaa*, jonka osuus yhteisesiintyjistä on semfinlex-lait-aineiston tulosten perusteella kasvanut vuoden 1998 jälkeen ajanjaksolla 1998–2009 ja verbin käyttö yhteisesiintyjänä yleistyy edelleen.

EU-lähtöinen yhdyssana *myyntitoimenpide*

Yhdyssana *myyntitoimenpide* toimii tässä alaluvussa esimerkkeinä EU-lähtöisistä yhdyssanoista. Näiden yhdyssanojen edusosana *toimenpide* säilyttää EU-teksteistä löytyneen merkityksen, joka on lähellä **säädöstä** tai **määräystä**. *Myynti*-alkuisia yhdyssanoja on EU-aineistossa 597 kappaletta ja verrannaisaineistossa (SL) 673 kappaletta. *Myyntitoimenpiteet koskevat tuotteita*, tai niihin *sovelletaan säännöksiä*, kun taas *polkumyyntitoimenpiteitä sovelletaan* tai *jätetään soveltamatta*, niillä on *soveltamisaika*, niitä *kierretään* tai ne voivat *olla voimassa*.

Muiden tutkittujen substantiivien joukosta EU-aineistosta löytyy yhdyssanoja, joiden määriteosana on *toimenpide* seuraavasti: *myynti* + *toimenpide* (4), *polku* + *myynti* + *toimenpide* (8). Kumpaakaan muotoa ei esiinny verrannaisaineistossa (SL).

Ainoa molemmissa aineistoissa esiintyvä substantiivi, jossa määriteosa on *myynti* ja edusosa jokin tutkituista substantiiveista, on *toiminta*. EU-aineistossa näitä esiintymiä on 6 ja SL-aineistossa 13 kappaletta. Esiintymien yhteydessä tutkimusikkunan laajuudessa kontekstissa esiintyy vain yksi verbi *harjoittaa*, molemmissa aineistoissa kuitenkin vain kerran.

Tutkimusaineistossa yhdyssanan määriteosa muuttaa tulosten perusteella edusosan sanakirjamerkitystä tai muuten tavallisinta merkitystä. Havainto vahvistaa johdopäätöstä, jonka mukaan tutkimuskohteilla on EU-kontekstissa myös EU-perusteinen merkitys.

Merkityksen muutos tai tarkentuminen: Muita havaintoja yhdyssanoista

Tutkin tarkemmin kymmentä molemmissa aineistoissa esiintyvää yhdyssanaa, jotka muodostavat tutkimusaineistossa yhteisesiintymiä tutkimusverbien *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä* ja *toteuttaa* sekä näiden verbikantaisten substantiivien kanssa. Tutkittujen yhdyssanojen määriteosat ovat *erityis*, *hoito*, *keskeyttämis*, *korjaus*, *lisä*, *suoja*, *tuki*, *turva*, *valvonta* ja *varo*. Tavoitteena on selvittää, ovatko havainnot verbikollokaattien esiintymisestä osana potentiaalista yhteisesiintymää tutkimuskohteen *toimenpide* kanssa samansuuntaisia kuin tutkimuskohteesta *toimenpide* itsenäisenä lekseiminä tehdyt havainnot. Tutkittavat määriteosat on valittu molemmille aineistoille yhteisestä materiaalista sen perusteella, että ne yhdessä edusosan *toimenpide* kanssa esiintyvät mahdollisimman usein sekä EU-aineistossa että SL-aineistossa. Esitän yhteenvedon näiden kymmenen yhdyssanan verbikollokaateista taulukossa 54. EU-aineistossa yhteisesiintymiä on huomattavasti enemmän ja niissä on enemmän vaihtelua kuin SL-aineistossa.

Taulukko 54. Kymmenen yhdyssanan, joissa *toimenpide* on määriteosana, tutkittujen verbikollokaattien esiintymistiheydet aineistoissa. (EU ja SL)

			<i>ryhtyä</i>	<i>soveltaa</i>	<i>suorittaa</i>	<i>säätää</i>	<i>tehdä</i>	<i>toteuttaa</i>
toimenpide	EU	2180	40	198	23	203	19	656
	SL	2763	970	5	300	17	68	90
erityis-	EU	52		3+7		1		5+2
	SL	10		1				2
hoito-	EU	9		5				
	SL	23			1			
keskeyttämis-	EU	4						2
	SL	10						4
korjaus-	EU	2						1
	SL	7	1		1			
lisä-	EU	12						10
	SL	1					1	
suoja-	EU	32		1				10+2
	SL	9						
tuki-	EU	34				2		3
	SL	4						
turva-	EU	10		2+4+1		1		1
	SL	3						
valvonta-	EU	52		4+3		7		6
	SL	13				1+1		
varo-	EU	6						6
	SL	9	1					

Esiintymistajuuksissa huomionarvoista on, että verbikollokaattien välillä on huomattavaa vaihtelua aineistojen välillä ja että SL-aineistossa tutkimusverbit esiintyvät yhdyssanojen yhteisesiintyjinä huomattavasti vähemmän tajaan kuin EU-aineistossa. Havaintoja SL-aineistossa on vain 14 verrattuna EU-aineiston 88 ha-

vaintoon.

Yhteisesiintymistä *erityistoimenpide*, *lisätoimenpide* ja *varotoimenpide* ovat SL-aineistossa yhdyssanoja, joissa edusosa *toimenpide* vaikuttaisi merkitykselliseltä lisältä yhdyssanan merkityksen rakentumisessa. Muista voisi intuitiivisesti ajatellen jättää edusosan pois ilman, että merkitys juuri muuttuisi. *Korjaustoimenpiteen* sijaan voitaisiin ryhtyä *korjaamaan*, ja *tuki-* tai *turvatoimenpiteen* saattaisi korvata *tuki* tai *turvaaminen*. Näyttäisi siltä, että *toimenpiteen* rooli säädöskielen yhdysosanmuodostuksessa on jonkinlainen tehostaminen, jolloin mahdollisesti koetaan kielen olevan jämäkämpää, tehokkaampaa tai vaikuttavampaa. Kuitenkin yhdyssanaa *turvatoimenpide* tarkemmin tarkasteltaessa EU-aineistosta löytyy seuraava mielenkiintoinen esiintymä, jossa yhdyssana noudattaa erillisenä esiintyvän hakusanan tulkintaa:

- (54) ja riittävät tiedot sopimuksen mukaisten turvatoimenpiteiden täytäntöönpanemiseksi. (EU) || Rivi 150670

Havaintojen perusteella voidaan todeta, että *turvatoimenpide* saa **säädöksen, lain** tai vastaavan entiteetin merkityspiirteitä. On mahdollista, että esimerkiksi *valvontatoimenpide* on lainsäädännössä tarkemmin määritelty käsite kuin **valvonta**. Silloin olisi kuitenkin todettava, että termi *toimenpide* ulottaa merkitystä muuttavaa vaikutusta määriteosaansa, eikä päinvastoin.

7.3 Semanttinen preferenssi ja semanttinen prosodia

Olen tarkastellut semanttista preferenssiä ja semanttista prosodiaa tutkimuksessani määrällisten havaintojen ohessa kirjaamalla konkordanssiriveiltä yhteisesiintymistä löytyneet aihealueet, mutta en ole niitä kyennyt aineiston laajuuden vuoksi käsittelemään määrällisesti tilastointia varten. Koska tutkimusverbien määrät varsinaisessa tutkimusaineistossa (EU) ja ensimmäisessä verrannaisaineistossa (SL) ovat lomittaisia siten, että toisessa aineistossa taajaan esiintyvä verbi esiintyy toisessa vähemmän taajaan ja päinvastoin, on käsityön määrä kaikkien verbien tarkastelussa huomattava. Niiden tavoitteena on avata keskustelua aiheesta ja pohtia jatkotutkimuksen tarvetta. Semanttisen prosodian ja semanttisen preferenssin tarkastelukulma tutkimuksessani on yhteisesiintymän V+*toimenpide* tekstiyhteydessä esiintyvien kielellisten elementtien kollokaattien tarkastelussa.

Semanttisen prosodian osoituksena pidetään esimerkiksi kieltosanan tai negatiivista merkitystä kantavan verbin esiintymistä vakiintujana. EU-aineistossa ainoastaan verbin *ryhtyä* tekstiyhteydestä löytyy selkeästi negatiivisia kielellisiä elementtejä. Yhteisesiintymän *ryhtyä* + *toimenpide* tekstikontekstissa esiintyy negatiiviseksi

koettavia verbejä, esimerkiksi verbit *ehkäistä* ja *estää* (mm. lainvastaisuuksia tai puutteellisuuksia) sekä *ratkaista* (mm. ongelmia). Negatiivisen konnotaation tuottavia substantiiveja ovat muun muassa *seuraamus*, *velvoite*, *epäkohta* ja *ongelma*. Suurin osa hakutuloksista on negatiivisia ja jäljelle jäävien muutamien esiintymien suhteen asia jää epäselväksi, koska tutkimusikkuna ei sisällä riittävästi tietoa tulokinnan tueksi.

SL-aineistossa osa konteksteista on myös samalla tavalla negatiivisesti värityneitä, jolloin konkordanssiriveiltä löytyy negatiiviseksi tulkittavia ilmauksia, esimerkiksi *puutteellisuus*, *epäkohta*, *vahingollinen vaikutus*, *palvelusrikkomus*, *ehkäiseminen*, *estäminen* ja *syytteen nostaminen*. Osa havainnoista on kuitenkin neutraaleja tai positiivisia, kuten esimerkiksi *tehtävän* tai *viran täyttäminen*, *hoidon järjestäminen*, *toiminta*, *hankkiminen* tai *selvittäminen*. EU-aineistossa vastaavassa kontekstissa esiintyy usein verbi *toteuttaa*, mutta toimenpiteitä myös *sovelletaan* (tietosuojaan turvaamiseksi) ja *suoritetaan* (*uudelleenlastaus*). Kääntäjä saattaa pyrkiä välttämään verbin *ryhtyä* valintaa vastineeksi negatiivisen semanttisen prosodian vuoksi.

Esimerkin 55 negatiivinen prosodia syntyy verbistä *estää*. (EU)

- (55) käyttävät immateriaalioikeuksia tai ryhtyvät muihin toimenpiteisiin estääkseen käyttäjiä tai jälleenmyyjiä hankkimasta (EU)

Yhteisesiintymän *soveltaa* + *toimenpide* tekstikontekstissa esiintyy samoin negatiiviseksi luokiteltavia lekseemejä, kuten *ei/eikä*, *lopetettava*, *tilapäinen*, *väliaikainen*, *ehkäisevä*, *rajoittaa* jalykätä. Pääosa havainnoista on kuitenkin neutraaliksi luokiteltavia, joten muiden neljän tutkimusverbin tavoin verbin negatiivisesta semanttisesta prosodiasta ei ole näyttöä.

Tulosten perusteella ainoastaan verbin *ryhtyä* tekstiyhteyden tarkastelu tuottaa hypoteesin semanttisen prosodian perusteella tapahtuvasta lähisyronymisen yhteisesiintymän valinnasta. Verbin *ryhtyä* semantiikkaa tutkimalla olisi mahdollista tarkistaa, onko saman kaltainen semanttinen prosodia havaittavissa yleisemminkin erikoiskiellisissä aineistoissa.

7.4 Assosiaatio ja vastavuoroisuus

Assosiaation ja vastavuoroisuuden tarkastelussa esitän joitakin huomioita yhteisesiintymien tekstiyhteydessä esiintyvien saneiden aihealueista ja kartoitan vastavuoroisuutta tarkastelevan HARSU/YLSU-näkökulman soveltuvuutta omaan tutkimuskohteeseeni ja aineistooni.

Havaintoja saneista ja aihealueista tutkimusverbien tekstiyhteydessä

Tutkimusverbien tekstiyhteydestä nousee joitakin usein toistuvia saneita, joista voi päätellä aihealueen. Aineiston perusteella ei voi suoraan päätellä, vaikuttaako aihealue verbin valintaan, koska havainnoista ei ole tehty kattavaa määrällistä tutkimusta. Useimmat verbeistä ovat niin tavallisia kaikissa tutkituissa aineistoissa, että niiden tarkempi tutkimus ei ole ollut tämän tutkimuksen puitteissa mahdollista.

Verbi *ryhtyä* liittyy molemmissa aineistossa usein **maatalouden, rakennusten perusparannuksen** ja **korjausrakentamisen** aihealueisiin.

Verbin *soveltaa* tekstiyhteydessä on säädöksiin liittyvää erikoissanastoa, kuten *laki, asetus, säännös, pykälä, momentti* ja *sopimus*. Yhteisesiintymän tekstiyhteydestä löytyy eniten **kaupapolitiikkaan** liittyvää sanastoa. Pelkästään adjektiivivi *kaupapoliittinen* esiintyy havaintojen joukossa 20 kertaa ja lisäksi aineistosta löytyvät samaa tarkoittavat variaatiot *kaupapolitiikkaa koskeva toimenpide, kaupapoliitiikan mukaisesti määrätty toimenpide* ja *kaupapoliitiikan toimenpide*. Konkordanssit sisältävät siten usein erilaisia tulleja, vientiä ja tavaroiden luovutusta koskevaa sanastoa. SL-aineistossa verbin *soveltaa* esiintyminen osana yhteisesiintymää *soveltaa + toimenpide* on niin vähäistä, ettei aineistosta voi muodostaa käsitystä tässä suhteessa. Vain yhdestä konkordanssista löytyy samaan aihealueeseen viittava sana *tavara*. LaDi-aineiston laeissa verbi *soveltaa* esiintyy lainkäytön kontekstissa ja EU-asetuksissa lisäksi yhteisöön liittyvässä (yhteisömalli). Direktiiveissä sen ala on näitä laajempi sisältäen myös yleisempiä asioita kuten *menettely* ja *vaatimus*.

SL-aineistossa verbi *suorittaa* on tavallinen yhteisesiintymän verbi aihealueilla **tulli** ja **rahoitus**. Esimerkkejä tekstiyhteydessä tällöin esiintyvistä saneista ovat *uhkasakko, rikkomus, rahoitusväline* ja *varat*. LaDi-aineistossa verbi *suorittaa* esiintyy laeissa enimmäkseen **taloudellisissa** tilannekonteksteissa (maksut), mutta myös yleisverbinä (valvonta), direktiiveissä verbi esiintyy niiden lisäksi myös muissa tekstiyhteyksissä, kuten tiedotus (palvelu, tiedottaminen) ja asetuksissa se näyttää toimivan yleisverbinä (käsittely, tutkinta, menettely).

Verbiä *säätää* käytetään EU-aineistossa lukuisten **komiteoiden** nimien (mm. *lääkevalmistekomitea, valmisteverokomitea, kalatalouden rakenteiden pysyvä hallintokomitea*) yhteydessä. Tekstiyhteydestä löytyy kyseisiin aloihin liittyvää yleistä sanastoa (esim. *maatalouspolitiikka, tulliviranomainen*). Komiteoiden nimet muodostavat kuitenkin pääosan tekstiyhteyden tulkinnassa. SL-aineistossa verbiä *säätää* esiintyy ennen kaikkea lain ja säädösten yhteydessä ja tekstiyhteydessä esiintyvät muun muassa *öljyvahinkojen torjuminen, tapaturmavakuutuslaki* ja *tupakkavalmisteet*.

Verbin *tehdä* kohteita EU-aineistossa ovat **päätökset, sopimukset, esitykset** ja **työ** ja sen tekstiyhteydessä on tyypillisesti **prosessiin** liittyvä ilmaus.

Verbin *toteuttaa* kohteita EU-aineistossa ovat *hanke*, *ohjelma*, *pyrkimys*, *päämäärä*, *suunnitelma*, *valvonta* ja *tehtävä* ja *yhteistyö*. LaDi-aineistossa kaikissa tekstilajeissa verbin *toteuttaa* tekstiyhteydestä löytyy tavoitteita ja hankkeita, joten se liittyy usein erilaisiin **prosesseihin**. Asetuksista löytyy lisäksi **kumppanuutta** ja **oikeustoimia** ja direktiiveissä **velvollisuuksia**, **vaatimuksia** ja **politiikkaa**.

EU-aineistossa verbiien kontekstista löytyy epätyypillisiä aihealueita, muun muassa verbi *säätää* esiintyy *poikkeuksen*, *tuen* ja *tuotantotuen* tekstiyhteydessä. Verbin *soveltaa* tekstiyhteydestä löytyy SL-aineistoon nähden epätyypillisiä termejä, muun muassa *poikeus* ja *tullittomuus* ja *valvonta*, *palvelua*, *kuljetusta* ja *tutkimusta* suoritetaan.

Tutkimuksen perusteella näyttäisi mahdolliselta, että yhteisesiintymän merkityksen määräytymisessä verbillä on EU-aineistossa pienempi rooli ja että verbiien yhteisesiintymien profiilit ovat EU-aineistossa joustavampia kuin SL-aineistossa. Havainnot ovat suuntaa-antavia ja alustavia, mutta ne tukevat verbiyhteisesiintymien ja tutkimuskohteen yhteisesiintymistä tekemiäni havaintoja.

Tutkimustuloksista nousee havainto tutkimuskohteen *toimenpide* merkityksestä termin *säädös* synonyyminä, jota tukee *toimenpiteen* yhteisesiintyminen tutkimusverbin *säätää* kanssa ja esiintymien tekstiyhteyden tutkiminen vahvistaa havaintoa. Vakiintuman *säätää* + *toimenpide* haku tutkimusaineistojen ulkopuolelta FinLexistä tuotti tulokseksi Hallituksen esityksen Eduskunnalle⁶³ vuodelta 2004. Esityksen tekstissä *toimenpiteellä* on säädöksille tyypillisiä piirteitä ja sen mukaisia verbikollokaatteja: esityksessä ehdotetaan, että Eduskunta *hyväksyy toimenpiteen*, että *toimenpiteellä perustetaan* sihteeristö, että *toimenpiteessä määrätään* muun muassa tehtävistä ja todetaan, että perustamisesta tehty *toimenpide* ei ole vielä *tullut* kansainvälisesti *voimaan*. Havainnon perusteella pidän selvänä, että *toimenpide* ei enää kymmeneen vuoteen ole ollut säädöskielessä käsite, jolla on kaikissa konteksteissa yhteneväinen merkityssisältö.

Havainto käytännön muuttumisesta (LaDi)

Mikä arvo on yksittäisillä havainnoilla? Havainto saattaa olla seurausta ainutkertaisesta tapahtumasta, mutta se on myös olemassa oleva ja tapahtunut ratkaisu. LaDi-aineistosta löytyy yksittäinen esimerkki, joka puoltaa hypoteesia verbin *toteuttaa* yleistymisestä ainakin tutkimuskohteina olevien substantiivien yhteydessä EU-asetusten määrän koko ajan lisääntyessä.

⁶³Hallituksen esitys Eduskunnalle Etelämannerta koskevan sopimuksen sihteeristön perustamisesta tehdyn toimenpiteen hyväksymisestä. <http://finlex.fi/fi/esitykset/he/2004/20040064>

Ensimmäinen esimerkki on vuoden 1996 ja toinen esimerkki vuoden 2001 direktiivistä. Verbi *suorittaa* on vaihtunut verbiin *toteuttaa* kontekstissa, joka näyttäisi olevan mahdollisimman identtinen; molemmissa **rajoitetaan käyttöalaa**.

Kuva 7.3: Vuoden 1996 direktiivistä, hakusana *toimenpide*

***	,	,	,	PUNCT	PUNCT COMMA	w2875
***	jäsenvaltion	jäsenvaltio	jäsen#valtio	N	N GEN SG	w2876
***	on	olla	olla	V	V COP PRES ACT SG3	w2877
***	suoritettava	suorittaa	suorittaa	PCP1	PCP1 PSS POS NOM SG	w2878
***	kaikki	kaikki	kaikki	PRON	PRON Q NOM	w2879
***	tarvittavat	tarvita	tarvita	PCP1	PCP1 PSS POS NOM PL	w2880
***	toimenpiteet	toimenpide	toimen#pide	N	N NOM PL	w2881
***	sen	se	se	PRON	PRON DEM GEN SG	w2882
***	käyttöalan	käyttöala	käyttö#ala	N	N GEN SG	w2883
***	rajoittamiseksi	rajoittaminen	rajoittaminen	N	N DV-MINEN TRA SG	w2884
***	,	,	,	PUNCT	PUNCT COMMA	w2885

Kuva 7.4: Vuoden 2001 direktiivistä, hakusana *toimenpide*

***	,	,	,	PUNCT	PUNCT COMMA	w4079
***	sen	se	se	PRON	PRON DEM GEN SG	w4080
***	on	olla	olla	V	V COP PRES ACT SG3	w4081
***	toteutettava	toteuttaa	toteuttaa	PCP1	PCP1 PSS POS NOM SG	w4082
***	kaikki	kaikki	kaikki	PRON	PRON Q NOM	w4083
***	tarvittavat	tarvita	tarvita	PCP1	PCP1 PSS POS NOM PL	w4084
***	toimenpiteet	toimenpide	toimen#pide	N	N NOM PL	w4085
***	osatekijän	osatekijä	osa#tekijä	N	N GEN SG	w4086
***	käyttöalan	käyttöala	käyttö#ala	N	N GEN SG	w4087
***	rajoittamiseksi	rajoittaminen	rajoittaminen	N	N DV-MINEN TRA SG	w4088
***	,	,	,	PUNCT	PUNCT COMMA	w4089

Merkitseviä ja tavallisia yhteisesiintymiä

HARSU (merkitsevä yhteisesiintymä) ja YLSU (tavallinen yhteisesiintymä) ovat tässä tutkimuksessa laskennallisia käsitteitä ja liittyvät siten yhteisesiintymien kartoitukseen, kun taas niiden rinnakkaistermit **merkittävä** ja **satunnainen** ovat merkitykseen sitoutuvia käsitteitä ja siten kuvaavat vakiintuman vastavuoroisuutta. Seuraavassa arvioidaan, miten tutkimusaineiston hakusanat ja tutkitut verbikollokaatit suhtautuvat näihin laskennallisiin käsitteisiin ja miten Sinclairin ajatus vastavuoroisuudesta näyttäisi toteutuvan tutkimusaineiston valossa (ks. alaluku 2.2.4).

Merkittävä yhteisesiintymä on Sinclairin harvinaisempaan kielelliseen elementtiin suuntautuva yhteisesiintymä (HARSU, jossa yhteisesiintyjä on aineistossa harvinaisempi kuin hakusana) ja tavallinen yhteisesiintymä on Sinclairin yleisempään elementtiin suuntautuva yhteisesiintymä (YLSU, jossa yhteisesiintyjä on aineistossa tavallisempi kuin hakusana).

Teoriassa tutkimusaineistossa HARSUja olisivat *ryhtyä + toimenpide*, *suorittaa + toimenpide* ja *ryhtyä + toimi*. Verrannaisaineistossa merkittäviä vakiintumia olisivat *ryhtyä + toimenpide* ja *toteuttaa + toimenpide*. Mainituissa vakiintumissa verbikolokaatti (vakiintuja) toimisi hakusanan merkitystä rajoittaen tai tarkentaen, mutta muut tutkitut kolokaatiot olisivat kompositionaalisempia eikä yhteisesiintymä niissä tapauksissa olisi semanttisesti motivoitu. Tutkimusaineistoissani yhteisesiintymien luokittelu merkitseviin ja tavallisiin yhteisesiintymiin tai merkittäviin ja satunnaisiin myötäesiintymiin ei tunnu intuitiivisesti toimivalta. Esimerkiksi *suorittaa + toimenpide* olisi EU-aineistossa merkitsevä, mutta SL-aineistossa satunnainen yhteisesiintymä. Mitään selkeää syytä näille eroille aineistoista ei löydy. Tutkimusverbien lukumäärät aineistoissa on esitetty taulukossa 55.

Jos hakusana olisi substantiivin sijaan verbi, jako HARSUihin ja YLSUihin toimisi päinvastoin, jolloin kompositionaalisesta yhteisesiintymästä tulisikin ei-kompositionaalinen merkitsevä vakiintuma. Koska yhteisesiintymän hakusana voi olla mikä tahansa ilmauksen kielellisistä elementeistä, tuntuu HARSU/YLSU-jaottelu toimivalta vain joissakin tietyissä konteksteissa, joissa hakusanan valinta ei ole vapaa (esimerkkinä Sinclairin tutkimat englannin fraasiverbit).

Esiintymien lukumäärä ei säädöskielessä korreloi suoraan tekstissä käsiteltävän lainalan tai ulkoisen maailmaan liittyvän aihealueen kanssa. Hakusanan ja yhteisesiintyjän esiintymistiheyksistä aineistossa suhteessa toisiinsa ei tämän tutkimuksen perusteella voi päätellä, onko kyseessä merkittävä vai satunnainen vakiintuma.

Johtopäätöksenä on esitettävä, että sinänsä mielenkiintoinen ja mahdollisesti tutkimisen arvoinen teoria on todennäköisesti tarkoitettu tietyssä kielellisessä kontekstissa toimivaksi eikä sellaisenaan sovellu tutkimuskohteen ominaispiirteen kartoitukseen tai kerro tutkitun yhteisesiintymän V+ *toimenpide* vastavuoroisuudesta. Tutkimukseni kannalta hyödyllistä oli tutkimusverbien lukumäärien selvittäminen EU-aineistossa ja SL-aineistossa.

Seuraavassa luvussa 8 tarkastelen tutkimustulosten valossa tarkemmin tutkimuskysymyksiä ja esittelen tutkimusaineiston perusteella tehdyt johtopäätökset. Avaan näkökulmia tutkimuksen hyödyllisyydestä kääntäjälle ja pohdin tulosten valossa aiheen mahdollisia jatkotutkimusaiheita.

Taulukko 55. *toimenpide* esimerkkinä HARSU/YLSU-laskentaperusteista.

toimen- pide SL	V	SL	SL	toimen- pide EU	V	EU	EU	V yhteis- esiinty- jänä
2763	ryhtyä	1542	HARSU	2180	ryhtyä	128	HARSU	40
2763	soveltaa	6627	YLSU	2180	soveltaa	6069	YLSU	198
2763	suorittaa	8788	YLSU	2180	suorittaa	1875	HARSU	23
2763	säätää	17707	YLSU	2180	säätää	3990	YLSU	206
2763	tehdä	9194	YLSU	2180	tehdä	4262	YLSU	19
2763	toteuttaa	695	HARSU	2180	toteuttaa	1633	HARSU	656

Luku 8

JOHTOPÄÄTÖKSIÄ JA NÄKÖKULMIA

Suomen liittyttyä 1990-luvun alkupuolella silloiseen Euroopan yhteisöön kohtasivat yhteisön suomennetut tekstit Suomessa kovaakin julkista kritiikkiä. Tuolloin heräsi myös kysymys siitä, voisiko kansalaisten kokemuksia tekstien outoudesta selittää uusien säädöksiä vaativien alojen lisäksi teksteissä käytetyn sanaston epätyypillisyydellä.

Tämän tutkimuksen tavoitteena on ollut löytää vastauksia johdantoluvussa 1 esitettyihin tutkimuskysymyksiin ja arvioida tutkimushypoteesin paikkansapitävyyttä aineiston valossa. Tutkimuskysymyksissä kysyin, onko Euroopan yhteisössä (nyk. Euroopan unioni) säädettyjen asetusten (EU-asetusten) suomentamisen seurauksena suomalaisen säädöskielen syntynyt epätyypillisiä tai uusia yhteisesiintymiä (tutkimuskysymys 1). Kysymyksenasettelu on mukavasti kaksiulotteinen: toisaalta voidaan pohtia kysymystä suomentamisen näkökulmasta ja toisaalta ilmiötä voidaan arvioida puhtaasti määrällisesti aineiston leksikaalista rakennetta tutkimalla. Selvitettävänä on, millaisilta EU-asetuksissa tutkittujen tutkimuskohteen ja kuuden tutkimusverbin muodostamat yhteisesiintymät näyttävät verrattuna Suomen säädöskokoelman vuonna 1998 voimassa olleisiin vastaavien osa-alueiden säädöksistä löytyviin yhteisesiintymiin. Tavoitteena on ollut pohtia, onko suomalainen säädöskieli muuttunut EU-asetusten kielten myötä yhteisöihin liittymisvaiheen aikana ja aineiston keräämisen jälkeen (tutkimuskysymys 2). Alun perin tutkimuskysymysten joukossa oli kysymys siitä, millaisia syitä oletettujen epätyypillisten esiintymien taustalla mahdollisesti olisi. Sen vuoksi teoriaosa sisältää merkitykseen ja assosiaatioon liittyviä osioita. Tutkimuksen aikana havaittiin, että tutkimusaineisto ei anna mahdollisuutta syiden selvittämiseen, mutta merkityksen ja assosiaation pohdinta ja niihin liittyvät havainnot on säilytetty osana tutkimusta.

Arvelin, että EU-asetusten kieli on tutkittujen leksikaalisten yhteisesiintymien osalta erilaista kuin autenttinen suomalainen säädöskieli. Oletuksena oli, että uusien käsitteiden lisäksi on syntynyt myös uusia tapoja ilmaista jo olemassa olevia käsitteitä. Määrälliset sekä rakenteeseen, merkitykseen ja assosiaatioon liittyvät havainnot ja pohdinnat löytyvät analyysiluvuista 6 ja 7. Tarkastelen seuraavassa alaluvussa kysymystä siitä, onko Suomen säädöskielen saapunut uudenlaisia yhteisesiintymiä Euroopan yhteisöjen asetusten myötä. Alaluku 8.1 on omistettu tutkimuskysymyksiin vastaamiselle ja tutkimusmenetelmän arvioinnille, alaluvussa 8.2 pohdin tutkimuksen merkitystä ja jatkotutkimuksen mahdollisuuksia.

Yhteisesiintymisen termistö on haastavaa riippumatta tutkijan vapaudesta määrittellä käyttämänsä käsitteet. Käytän termiä kollokaatio yläkäsitteenä ja sen alalajeista termejä yhteisesiintyminen, yhteisesiintymä ja yhteisesiintyjä milloin kyseessä on määrällinen pohdinta ja termejä vakiintuminen ja vakiintuma, kun huomio on yhteisesiintymän merkitystä muodostavissa ominaispiirteissä. Yhteisesiintymä ja vakiintuma ovat ikkunoita samaan ilmiöön. Kuvaan tutkimuksen myötä syntynyttä hahmotelmaa kollokaatiosta ilmiönä tarkemmin alaluvussa 8.3.

Käsittelen tutkimuskysymyksiä niiden havaintojen perustella, jotka olen tehnyt tutkimuksen kohteiksi valittujen kielellisten elementtien esiintymistä varsinaisessa tutkimusaineistossa (EU) ja ensimmäisessä verrannaisaineistossa (SL). Vertaan havaintoja toiseen verrannaisaineistoon (semfinlex-lait), kolmanteen verrannaisaineistoon (LaDi) ja neljänteen verrannaisaineistoon (lehdet-osa). Havainnot on raportoitu luvuissa 6 ja 7. Kollokaation teoreettinen tarkastelu löytyy luvusta 2, aineisto on kuvattu luvussa 4 ja menetelmät luvussa 5.

8.1 Tutkimuskysymykset ja tutkimusmenetelmät

8.1.1 Tutkimuskysymyksiin vastaaminen

Tutkimuksen ensisijaisena tavoitteena oli vastata ensimmäiseen tutkimuskysymyksen: onko EU-asetusten kieli yhteisesiintymien valossa erilaista kuin suomeksi kirjoitettu säädöskieli? Toisen tutkimuskysymys käsitteli mahdollisuutta arvioida havaintujen epätyypillisten kollokaatioiden perusteella arvioida niiden vaikutusta Suomen säädöskielen kehitykseen. Esitän tässä keskeiset määrälliset ja rakenteeseen liittyvät havainnot vastauksena ensimmäiseen tutkimuskysymykseen.

Määrälliset ja rakenteeseen liittyvät havainnot

Tutkimus vastaa ensimmäiseen tutkimuskysymykseen aineistosta tehtyjen havaintojen pohjalta varsin kattavasti. Kysymystä on tarkasteltu yhteisesiintymien määrällisellä ja rakenteellisella kartoituksella sekä katsauksella merkitykseen liittyviin ominaispiirteisiin. Työn laajuudesta johtuen merkitykseen liittyvien havaintojen osuus on määrällisiä suppeampi, mutta niidenkin tulokset synnyttävät uusia hypoteeseja jatkotutkimuksia varten. Kollokaation ominaispiirteistä ja erityisesti ilmiön laajuudesta tutkimus tarjoaa uudenlaista tietoa rakenteeseen ja syntaksiin liittyen, mutta ei vastaa kysymyksiin ilmiön syntymekanismeista tai merkityksen muodostumisesta.

Tutkimuskohde *toimenpide* on tavallisempi varsinaisessa tutkimusaineistossa (EU) kuin ensimmäisessä verrannaisaineistossa (SL). EU-aineistossa esiintymiä on 2 180, kun aineiston koko on 1 369 299 sanetta ja esiintymiä SL-aineistossa on 2 763 per 3 559 229 sanetta

Yhteisesiintyjien *ryhtyä*, *säätää*, *soveltaa*, *suorittaa*, *tehdä* tai *toteuttaa* jakamat varsinaisessa tutkimusaineistossa (EU) ja ensimmäisessä verrannaisaineistossa (SL) poikkeavat toisistaan. Tavallisin verbi termin *toimenpide* lähiympäristössä EU-aineistossa tutkittujen verbien *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä* ja *toteuttaa* joukossa on *toteuttaa*. Sen osuus tutkittujen verbien esiintymistä *toimenpiteen* kanssa on lähes 60 prosenttia, kun taas sen osuus verrannaisaineistossa (SL) on vastaavasti vain reilu kuusi prosenttia. Sen lisäksi taajaan esiintyviä verbikollokaatteja EU-aineistossa ovat *säätää* ja *soveltaa*, joilla molemmilla on noin 18 prosentin osuus kaikkien kuuden tutkitun verbikollokaatin esiintymistä yhteisesiintymän V + *toimenpide* komponenttina suhteessa muihin tutkittuihin verbeihin.

Sekä *säätää* että *soveltaa* ovat verrannaisaineistossa (SL) erittäin harvinaisia, kun 1450 tutkitun yhteisesiintymän V + *toimenpide* joukossa *säätää* esiintyy yhteisesiintymän osana vain 17 kertaa ja verbi *soveltaa* ainoastaan viisi kertaa. EU-aineistossa vastaavat luvut per 1 142 tutkittua yhteisesiintymää ovat 206 ja 198.

SL-aineiston tavallisimmat *toimenpiteen* yhteisesiintyjät ovat *ryhtyä* 66 prosentin osuudella ja *suorittaa* 20 prosentin osuudella kuuden tutkitun yhteisesiintyjän esiintymistä osana yhteisesiintymää V + *toimenpide*. EU-asetuksissa *toimenpiteitä* **toteutetaan**, **säädetään** ja **sovelletaan**, kun taas Suomen säädösteksteissä niihin **ryhdytään** ja niitä **suoritetaan**. Tutkimustulos vahvistaa myös havainnon *soveltaa*-verbin epätyypillisestä käytöstä varsinaisessa tutkimusaineistossa (EU) ja verbin *suorittaa* aliedustuksesta *toimenpiteen* yhteydessä EU-asetuksissa verrattuna SL-aineistoon sekä havainnon *säätää*-verbin epätyypillisestä käytöstä *toimenpiteen* yhteydessä osana tutkittua yhteisesiintymää. Tarkastelen tarkemmin merkitykseen liittyviä havaintoja verbien **suorittaa**, **säätää** ja **toteuttaa** esiintymistä yhteisesiintyjinä seuraavassa osiossa.

Termin *toimenpide* yhteyteen valikoituvat eri verbit EU-aineistossa verrattuna SL-aineistoon. Näin selvä ero termin *toimenpide* tutkittujen yhteisesiintyjien jakauksessa puoltaa ajatusta EU-kielen rantautumisesta suomalaiseen säädöskieleen. Mikään verbeistä ei täysin puutu havaintojen joukosta. On aihetta epäillä, että EU-aineistossa tyypilliset *säätää* ja *soveltaa* olisivat saapuneet SL-aineistoon esimerkiksi Euroopan yhteistyöhön liittyvien, mahdollisesti vuonna 1949 perustetun Euroopan neuvoston toimintaan liittyvien käännostekstien mukana.

Yhteisesiintymän osana hakusana *toimenpide* on tutkituissa yhteisesiintymissä sekä EU-asetuksissa että Suomen lakiteksteissä lähes aina monikossa silloin, kun yhteisesiintyjänä on jokin verbeistä *ryhtyä*, *soveltaa*, *tehdä* tai *toteuttaa*. Poikkeuksen muodostavat siten verbikollokaatti *suorittaa*, jonka kanssa esiintyessään EU-aineistossa *toimenpide* on yksikössä noin viidesosassa havainnoista ja verrannaisaineistossa (SL) vastaavassa rakenteessa lähes 70-prosenttisesti yksikössä. Kun kääntäjä valitsee vastineeksi termin *toimenpide* yksikössä olevan muodon, on todennäköistä, että hän valitsee sen yhteisesiintyjäksi verbin *suorittaa*. Havainnon tueksi tarvittaisiin syventävä tutkimus yhteisesiintymän nyt tutkituilla aineistoilla sekä yleisemminkin yleiskielisellä ja muiden erikoiskielisten tekstilajien aineistoilla.

Yhteisesiintymän muodostavien kielellisten komponenttien sijaitseminen jonossa tiettyssä positiossa ei osoittautunut yhteisesiintymisen muodostumisen kannalta yhdenmukaiseksi. Tutkitut verbit profiloituvat eri tavoin suosimaan eri positioita, mikä selittyy osin verbikantaisten substantiivien osuuksilla, verbin passiivimuodoilla (joita ei ole tässä tutkimuksessa erikseen tarkasteltu), jotka kansoittavat hakusanan jälkeisiä positioita ja osin vakiolausekkeilla, jotka puolestaan selittävät esiintymiä kauempana hakusanasta olevissa positioissa. Tulokset antavat aihetta niiden tarkempaan tarkasteluun ja luokitteluun, johon käsillä oleva tutkimus vasta tuottaa aineksia.

Tutkimuksen havainnot tutkimusikkunan ulkopuolelta löytyvistä yhteisesiintyjistä perustuvat LaDi-aineiston sisältämien direktiivien tutkimukseen, jotka edustavat pientä otosta EU-kielestä. Ilmiötä virke- ja kappalerajoja ylittävänä mekanismina olisi syytä tarkastella kattavammalla dependenssijäsennetyllä aineistolla. Tarkasteltava ilmiö on osoittautunut niin moniulotteiseksi, että tämän tutkimuksen anti on tässäkin kohtaa avata ikkunoita tulevalle tutkimukselle.

Merkitykseen ja assosiaatioon liittyvät havainnot

Yleiskieltä edustavassa sanomalehti- ja aikakauslehtiaineistossa verbi *tehdä* esiintyy taaajimmin tutkituissa yhteisesiintymissä. Se on kaikissa säädöskieltä sisältävissä aineistoissa epätavallinen lekseemin *toimenpide* yhteisesiintyjä. Neljännen verrannaisaineiston verbien *ryhtyä*, *soveltaa* ja *säätää* jakaumat ovat lähellä SL-aineistolle ja verbi *toteuttaa* EU-aineistolle tyypillistä jakaumaa.

Säädöskielessä suhteellisen harvat lekseemit tuottavat alalle tarpeellisen sisällön. Tosin oikeuskieli laajemmin ja säädöskieli oikeuskielen tekstien kautta jakautuvat eri elämänaloja kattaviin luokkiin, joissa voidaan olettaa olevan oma tyypillinen sanastonsa. Tässä tutkimuksessa sekä tutkimusaineisto että sen verrannaisaineisto on koostettu teksteistä, joissa pääpaino on taloudellisessa yhteistyössä ja sanasto painottuu sen mukaisesti talouden käsitteisiin.

Tutkimusaineiston perusteella lekseemin *toimenpide* tarkoite EU-kontekstissa poikkeaa suomalaisessa säädöskielessä käytetyn termin tarkoitteesta. Termi *toimenpide* on saanut EU-tekstien myötä uuden merkityksen. Tätä tulkintaa tukevat erityisesti havainnot verbien *säätää* ja *soveltaa* esiintymistä verbiyhteisesiintyjien joukossa. Suomen lainsäädännössä *lakeja* ja *asetuksia* säädetään ja *määräyksiä* sovelletaan, mutta EU-kontekstissa myös *toimenpiteitä* on asetusten kääntämisen aikoihin *säädetty* tai *sovellettu*.

Termillä *toimenpide* on EU:n IATE-termipankissa seitsemän toisistaan poikkeavaa englanninkielistä käännösvastinetta, muun muassa termit *act* ja *undertaking*, joiden merkitys liittyy säädöksiin tai lainsäädäntöön. Niiden takaisinkäännökset suomeen ovat *laki* ja *säädös*. Lekseemillä *toimenpide* on yhteisesiintymässä yhdessä verbikollokaatin *säätää* kanssa EU-kontekstissa myös *lakiin* tai *säädökseen* viitettava merkitys.

Yhteisesiintyjän vaikutuksesta tapahtuvasta merkityksen eriytymisestä esitän seuraavassa kokoavat havainnot kolmesta tutkitusta yhteisesiintymästä, joiden kohdalla tutkimustulokset ovat riittäviä osoittamaan merkityksen muutosta. Yhteisesiintyjät ovat *suorittaa*, *säätää* ja *toteuttaa*.

suorittaa + toimenpide

Yhteisesiintymä *suorittaa + toimenpide* sijoittuu usein merkityksensä puolesta **tullin** ja **rahoituksen** aloille. SL-aineistossa tulliin ja rahoitukseen liittyvää EU-asetuksille tyypillisestä sanastosta (mm. *rikkomus*, *uhkasakko*, *varat*) löytyy vain muutamia yksittäisiä esiintymiä tutkimusikkunan sisällä. SL-aineistossa suoritettavat *toimenpiteet* ovat monelta eri elämänalueelta, ja *suorittaa* vaikuttaa näin neutraalimmalta verbiltä kuin mitä sen tehtävä EU-kontekstissa antaa ymmärtää.

säätää + toimenpide

Substantiivi *toimenpide* ei sanakirjamääritelmien mukaan merkitse säädöstä, määrystä eikä lainkohtaa. Silti se muodostaa yhteisesiintymän verbin *säätää* kanssa EU-asetusten aineistossa. *Säätää* esiintyy muutoin molemmissa aineistoissa lainsäädäntöön liittyvissä konteksteissa (lukuun ottamatta EU-aineiston alla taulukos-

sa 55 mainittuja *poikkeusta* ja *tuotantotukea*), joten yhteisesiintymän *säätää + toimenpide* merkityksen muotoutumisessa substantiivin *toimenpide* merkityskenttä muuttuu sisältämään piirteitä, jotka liittyvät **säädöksiin, määräyksiin** ja muihin käsitteisiin, joita lainsäädännössä voidaan *säätää*. Mikäli näin ei ole, on pääteltävä, että käänösprosessissa on tapahtunut taivutusparadigman muutos; sen sijaan, että *säädetään toimenpiteestä, säädetään toimenpide*. Verrannaisaineistossa ei ole kyseisiä esiintymiä.

toteuttaa + toimenpide

Toimenpiteen toteuttaminen ei ole vanhastaan kuulunut Suomen säädöskielen idiomattisten yhteisesiintymien joukkoon säädöskielen ammattikielessä. Sen sijaan *toimenpiteeseen* ja useimmiten *toimenpiteisiin* on perinteisesti *ryhdytty*.

Tulosten perusteella *toimenpide* on molemmissa aineistoissa tavallinen termi, joka on uusien yhdyssanojen muodostamisessa aktiivinen ja joka osoittautuu selkeästi potentiaalisiksi vaihtoehdoksi useille lähtökielen termeille IATE-termipankin käänösvastineiden variaation perusteella. Lekseemin *toimenpide* luonne EU-terminä eroaa kuitenkin Suomen säädöskokoelman lakitekstien konventioista, joten Euroopan unioni on laajentanut termin merkityskenttää mahdollistamalla uusia kollokaatiorkenteita. Samalla kotoperäiset konventionaaliset odotukset termin käyttöympäristön suhteen kokevat muutoksen, kun uusia mahdollisuuksia avautuu. Mahdollista on, että termi profiloituu EU-termiksi, jolloin syntyy EU-toimenpide vanhan SL-toimenpiteen rinnalle ja niiden erilainen merkitys avautuu erikoiskielen ja ammattikielen käyttäjille kontekstinsa pohjalta.

Muutos Suomen säädöskielessä

Toiseen tutkimuskysymykseen pyrin vastaamaan lisäämällä tutkimusaineistoihin toisen verrannaisaineiston, lakien osuuden Finlex-tietokannan alkuperäissäädöksistä. Tutkin aineistoa semfinlex-lait kolmelta eri ajanjaksolta, joista ensimmäinen kattaa säädökset vuodesta 1960 alkaen vuoteen 1998, jolloin varsinainen tutkimusaineisto koottiin. Toinen ja kolmas ajanjakso kattavat kaksi seuraavaa vuosikymmentä. Aineiston valossa voidaan sanoa EU:n konventioiden näkyvän muutoksina termin käytössä ja Suomen säädöskokoelman kieli lähenee näiltä osin EU-asetusten kieltä.

Aineiston semfinlex-lait analyysin tulos vahvistaa havainnot varsinaisen tutkimusaineiston (EU) ja ensimmäisen verrannaisaineiston (SL) sisältämistä yhteisesiintymistä *ryhtyä + toimenpide*. Se oli ajanjaksolla 1960–1998 SL-aineiston tavoin selkeästi taajimmin esiintyvä yhteisesiintyjä. Verbin *ryhtyä* rooli yhteisesiintyjänä on kuitenkin vähentynyt tasaisesti vuoden 1998 jälkeen. Koska saneen *toimenpide*

lukumäärät säädöskielessä ovat vuoden 1998 jälkeen kasvaneet tasaisesti, voidaan olettaa, että verbin *ryhtyä* tilalle on tullut ja tulossa muita yhteisesiintyjä. Verbin *toteuttaa* lukumäärä aineistossa on kaksinkertaistunut kahden ensimmäisen vuosikymmenen kuluessa ja vuosina 2010–2018 sen käyttö on lisääntynyt yli kolmanneksella verrattuna edelliseen vuosikymmeneen 1999–2009. Sen esiintymistajuus *toimenpiteen* yhteisesiintyjänä on kasvanut lähes viisinkertaiseksi vuosina 1999–2009 verrattuna edelliseen vuosikymmeneen.

Ennen semfilex-aineiston julkaisua pyrin löytämään vastauksia kysymykseen Kielipankissa julkaistun, alunperin Kotuksessa kootun aineiston Lakeja ja Direktiivejä avulla, mutta se osoittautui rakenteensa ja tekstien alkuperän vuoksi soveltumattomaksi muutoksen jälkien havaitsemiseen. Sen sijaan pääsin aineiston avulla kartoittamaan esiintymiä direktiiveistä ja tarkastelemaan aineiston sisältämien asetusten tutkimusikkunan ulkopuolelle ulottuvaa tekstikontekstia. Vaikka direktiivit eivät ole osa varsinaista tutkimusaineistoa, voidaan niiden avulla hahmottaa, olisiko tutkittava yhteisesiintymä löytänyt tiensä suomalaisen säädöskielen myös niiden kautta. Lakeja ja Direktiivejä sisältävän aineiston osalta voidaan todeta, että sen EU-tekstit noudattavat EU-aineiston jakaumaa ja lait SL-aineiston jakaumaa. Yhteisesiintymiä oli eniten direktiiveissä ja niissä lähes yksinomaan verbi *toteuttaa*. Voidaan todeta myös tämän aineiston tukevan tutkimustulosta säädöskielen muuttumisesta.

Kollokaation käsite ja ominaispiirteet

Kollokaation käsitteen tarkentuminen ei ollut varsinainen tutkimuskysymys, mutta havaintojen perusteella voidaan todeta ilmiön olevan haastava ja hankalasti rajattavissa. Tutkimuksen myötä syntyneitä ajatuksiani olen kuvaan aikaisemmin tässä luvussa. Tässä osiossa otan kantaa tutkimuksen alkuvaiheessa tehtyihin oletuksiin kollokaation ominaispiirteistä. Ne on kuvattu luvussa 2.1.

Tutkimukseni sitoutuu vahvasti kontekstuaaliseen näkemykseen kielestä Tulokset eivät ole yleistettävissä, enkä näe tutkimukseni asettuvan sen paremmin puolustamaan kuin vastustamaan teoriaa kääntämisen universaaleista. Arvelen kuitenkin seuraavani Firthin jalanjalkia, kun totean, että nämä kaksi lähestymistapaa eivät helposti kohtaa toisiaan eikä oma tutkimukseni kysymykseen kääntämisen universaaleista lisää tietoa. Tutkimukseni tulosten valossa näen, että vaikka kollokaatio ilmiönä on kaikkialla, se ei ole samalla tavoin merkityksellinen kaikissa konteksteissa, tekstilajeissa ja tekstityypeissä. Säädöskielen kääntäjälle kontekstuaalisuus asettaa vaatimuksia sekä tilannekontekstin että tekstikontekstin hallitsemisesta, mutta merkitystä luovalle, ei-kompositionaaliselle vakiintumalle ei säädöskielessä ole funktiota. Katson, että kollokaatio säädöskielessä säilyttää kompositionaalisuutensa. Sen sijaan merkityksen hienovarainen muutos, joka heijastuu näkyviin yhteisesiintyjien

tarkastelun kautta, on kääntäjälle arvokas tieto.

Tutkimukseni asettuu kollokaatiotutkimuksen haaraan, jossa ilmiötä tarkastellaan ensisijaisesti käännöstieteen tarpeisiin. Käsitteen jako distributionaaliseen ja intensionaaliseen lähestymistapaan palvelee kääntäjää parhaiten silloin, kun molemmat ikkunat näihin ulottuvuuksiin ovat auki yhtä aikaa. Tutkimustulosten perusteella en ole kyennyt asettamaan raja-arvoja yhteisesiintymän havainnoille, joita tilastolliset menetelmät arvioivat. Kysymys siitä, missä kohtaa yhteisesiintymä muuttuu mielenkiintoiseksi kääntäjän kannalta on eri kysymys kuin ilmiön käsitteellinen määrittely. Kääntäjälle puhdas numeerinen tieto on toivottavasti mielenkiintoista. Kun havaitaan, että *toimenpide* on paljon käytetty termi erilaisten lähtökielen termien vastineena, voimme lähteä kartoittamaan sen kenttää ja tutkia, miten eri tavoin ja millaisista kielellisistä vaihtoehdoista se valitsee elementtejä yhteisesiintymään kohdekielessä. Kääntäjä voi tietoisesti valita kotouttavan tai vieraannuttavan käännösstrategian tai pohtia adekvaattisen ja hyväksyttävän strategian vaikutuksia kohdekielessä jo vakiintuneisiin konventioihin. Sinclairin tavoin päädyn ajattelemaan, että kahdesti löytyvä yhteisesiintymä on (kääntäjän kannalta) olemassa. Distributionaalinen näkemys säädöskielestä kattaa havainnot siitä, mitä on olemassa, millaista variaatiota termi itsessään tai sen käyttö sallii. Havaintojen päälle rakentuvat termin ja yhteisesiintyjän merkitystä tuottavat tasot, jotka voivat olla syntaktisia tai semanttisia ja jotka voivat vakiintua esimerkiksi leksikaalisen virittämisen tai virittymisen keinoin.

Painotus distributionaaliseen lähestymistapaan Evert (2005) ei tarkoita intensionaalisen poissulkemista. Kun yhteisesiintymiä on tunnistettu, niiden merkityksen tutkimisessa voidaan hyödyntää semanttisen preferenssin ja semanttisen prosodian tai assosiaatiota mittaavia menetelmiä. Tämän tutkimuksen yritykset ottaa kantaa semanttiseen prosodiaan jäävät negatiivisuuden rekisteröinnin tasolle intuitiivisen tiedon pohjalta ilman esiintymien määrällistä tarkastelua. On myös huomattava, että negatiivisuus on vain yksi semanttisen prosodian ilmenemismuodoista; hyvin voisi tutkia positiivisuuden ilmentymiä tai muita indikaattoreita kirjoittajan tai puhujan asenteista. Jos voidaan esimerkiksi todeta, että verbi *ryhtyä* esiintyy yleiskielessä aina negatiivisessa kontekstissa, voi sen avulla selittää kääntäjän mahdollisesti alitajuisesti välttävän verbin käyttöä kaikissa konteksteissa Kübler ja Volanschi (2012). Havaittiin, että EU-aineistossa verbi *ryhtyä* kyllä esiintyy negatiivisiksi luokiteltavien saneiden kanssa EU-aineistossa, mutta vastaavaa tutkimusta ei yleiskielisestä aineistosta voitu tämän tutkimuksen puitteissa tuottaa. Lisäksi kaikkien verbien tekstikonkreetista löytyi negatiivisiksi katsottavia sanoja, jolloin tarkempi arvio edellyttäisi sentimenttianalyysityökalua tai muuta ulkoista negatiivisuuden kriteeriä. Havainnot on kuitenkin jätetty tekstiin lukijan arvioitaviksi. Semanttinen preferenssi edellyttäisi samoin oman tutkimuksensa, jossa aihealueet kartoitettaisiin tarkemmin, jolloin niiden avulla olisi mahdollista päästä tarkastelemaan verbien valintoihin johtaneita

syitä. Arvelin, että assosiaatiota voisi mitata seuraten Sinclairin ajatusta siitä, että on merkityksellistä, kumpi yhteisesiintyjän komponenteista on hakusana – toisin sanoen, ovatko ne vaihdettavissa keskenään ja tuottuisiko samanlaisia kollokaatioita, jos roolit vaihdettaisiin. Tässä kompastuin aineistoni laatuun ja yhteisesiintyjien valintaan. Verbin ja substantiivin muodostama yhteisesiintyjä ei ole optimaalinen väline hypoteesin tutkimiseen.

Kysymys fraseologian ja terminologian rajanvedosta ei osoittaudu aineiston valossa merkitykselliseksi kysymykseksi. Vastoin Hausmannin Hausmann (2004) ajatuksia määrittelen tutkimuskohteeni termiksi ja yhteisesiintymän hakusanaksi ja vastoin Heidin Heid (2001) näkemyksiä olen valinnut yhteisesiintyjäksi erikoiskielissäkin taajaan käytettyjä verbejä, kuten *tehdä* tai *säätää*. Intensionaalisen näkökulman valitseminen työn pohjalle toisi kysymyksen paremmin esiin tutkimuskysymysten valinnan kautta. Tämä työ odottaa jatkossa tekijäänsä. Fraseologinen näkökulma edellyttää tavallisesti lähtökohtaisesti ei-kompositionaalisuuden kriteerin täyttymistä (ks. kollokaation käsite luvussa 2). Tutkimukseni yhteisesiintymät eivät toteuta tätä ehtoa, eikä tutkimiani yhteisesiintymiä näin voida pitää fraseologisina yksikköinä.

Olen tässä tutkimuksessa tarkistanut kollokaation muotoon ja ominaispiirteisiin liittyviä lähtöoletuksia. Ilmiönä kollokaatio ottaa erilaisia muotoja kieliaineuksessa ja sen laaja-alainen ymmärtäminen ja jäsentyminen kääntäjälle hyödylliseksi työkaluksi edellyttää luopumista muotoon liittyvistä kriteereistä. Distributionaaliset näkymät tarjoavat kysymyksiä herättäviä aineksia, kuten tutkimuksessani tutkimuskohteen kollokaattikandidaattien valinnassa kävi ilmi. Odottamattomat distributionaaliset havainnot tuottavat hedelmällisiä tuloksia. Merkitykseen ja assosiaatioon linkittyvät näkymät vakiintumaan puolestaan johdattavat kääntäjän kannalta mielenkiintoisille ajatuspoluille, jossa esimerkiksi omassa tutkimuksessani yhdyssanat ja verbikantaiset substantiivit kantavat yhtä lailla vakiintuman yhteistä merkitystä.

Semanttisen prosodian ja semanttisen preferenssin osalta tutkimukseni ei tarjoa juuri aihealueiden pinnallista hahmottamista syvempää tietoa. Vakiintuman kannalta aihealueiden kontrastointi tai semanttista prosodiaa rakentavien elementtien tarkastelu toisi havaintoihin uuden kerroksen, joka yhdistettynä distributionaaliseen tarkasteluun vasta kattaisi ilmiön laajuutta ja hyödyllisyyttä kääntäjälle ja terminpankkien kehittäjille.

8.1.2 Tutkimusmenetelmien arviointia

Kollokaatio on käsitteenä yhden tutkimuksen kohteeksi laaja ja monisyinen ilmiö. Sen tutkiminen erikoiskielten aineistolla osoittautui haastavaksi tehtäväksi ja tuloksena onkin enemmän kysymyksiä kuin vastauksia. Ilmiön kuvauksessa hahmottuneet

tavat tarkastella aihetta autenttisen tekstin läpi tuottivat parhaimmillaan ideoita ilmiön rajaamiseksi, mutta käytännön tutkimuksen puolella tutkimustulosten valossa kollokaation käsite ei juurikaan kirkastunut. Yhdyssanan olemus yhteisesiintyjänä oli tässä suhteessa yllätys ja yhdyssanojen tutkimusta yhdenlaisena kollokaation muotona olisi hyvä jatkaa.

Päätös pitää tutkimus vahvasti distributionaalisen tutkimuksen puolella on perusteltu aineiston laajuus huomioiden riippumatta Firthin näkemyksestä merkityksen tutkimuksen ensisijaisuudesta. Merkitykseen ja assosiaatioon liittyvät havainnot jäivät tutkimuksessa toissijaiseen asemaan ja vaikka se oli suunniteltua, on tässä kohtaa kritiikki paikallaan: vähemmällä laskennallisella työllä olisi mahdollisesti päästy samaan tulokseen määrällisten havaintojen puolella, jolloin tilaa olisi jäänyt enemmän merkityksen, fraseologian ja assosiaatioon liittyvien ominaispiirteiden tarkasteluun.

Tutkimusmenetelmä on yksinkertainen: lasketaan tutkimuksen kohteena olevat substantiivit ja niiden kanssa esiintyvät verbit molemmissa aineistoissa ja verrataan esiintymistaajuuksia keskenään käyttäen tilastollisia menetelmiä yhteisesiintymien tilastollisen merkitsevyyden mittareina. Käytännössä tarvitsin tilastollista merkitsevyyttä mittaavia menetelmiä vain tutkimuksen alkuvaiheessa, kun oli löydettävä kohtuullisen objektiivinen mittari tutkittavien substantiivien ja niiden yhteisesiintyjäehdokkaiden valintaan siten, että esiintymiä olisi riittävästi muttei kuitenkaan ylivoimaisen paljon laadullisen analyysin läpiviemiseksi. Lopulta tilastollisia testejä tarvittiin kuitenkin frekvensseihin pohjautuvan päättelyn validoimiseksi.

Olen määritellyt tutkimuskohteen tässä tutkimuksessa perusmuodon (lemma) sekä merkityksen perusteella (verbikantaiset substantiivit), saneen ortografisen muodon sijaan. Sinclairin tutkimuksissa lähtökohtana on usein muoto, jolloin voidaan erottaa helpommin homonyymiset esiintymät toisistaan ja ottaa huomioon kielipillisiä näkökantoja (Sinclair et al. (2004)). On mahdollista, että pintamuoto vaikuttaa kielellisen yksikön kollokaatioprofiiliin, mutta pintamuotopohjaisen kartoituksen tekeminen suomenkielisestä aineistosta on huomattavasti työläämpää ja vaatii isomman aineiston kuin englanninkielisestä aineistosta. Sinclair on ehkä oikeassa hahmottaessaan erilaisia kehyksiä, joiden sisällä myös vakiintuman komponenttien muoto vaikuttaa sen merkitykseen, mutta tämän tutkimuksen puitteisiin ei asian tarkastelu ole mahtunut. Olen tuottanut konkordanssit perusmuodoista, mutta tarkastellut tuloksena olevia konkordansseja pintamuotoisesta materiaalista, koska suomenkielistä perusmuotoistetusta tekstistä ei saa selville, onko hakusanan ja tutkimusverbin välissä dependenssisuhdetta.

Tutkimuksen aloittamisen jälkeen vuonna 2000 kieliteknologisten metodien arvioinnin ja kehittämisen saralla on tapahtunut paljon ja uusia työkaluja on saatu kieliaineistojen tutkimukseen. Oman tutkimukseni varsinaisen tutkimusaineiston (EU)

ja ensimmäisen verrannaisaineiston (SL) tutkimuksessa käytetty menetelmä on kuitenkin monilta osin ollut sidottu tutkimuksen alkuvaiheessa tehtyihin ratkaisuihin. Se on tavallaan hyväkin asia: menetelmä on pysynyt yksinkertaisena, eivätkä tekstikorpusten käsittelyn Unix-pohjaiset ohjelmat ja komentoperheet ole muuttuneet vuosien kuluessa. Sen ansiosta olen voinut tehdä uudelleen lähes kaikki ajot viimeksi kuluneiden muutaman vuoden aikana ja tulosten tarkistaminen on ollut pitkin matkaa mahdollista. Kaikki ohjelmat ja ohjelmistot ovat tutkimuskäyttöön vapaasti saatavilla. Toisen, kolmannen ja neljännen verrannaisaineiston osalta tutkimuksen ovat mahdollistaneet Kielipankin kautta käytettävissä olevat aineistot ja konkordanssityökalu Korp. Niidenkin osalta aineistot olen kuitenkin arvioinut ja laskenut esiintymät konkordanssiriveiltä manuaalisesti. Dependenssijäsennetyin aineiston saatavuus Kielipankissa jatkossa edistää sen julkaisemien kieliresurssien käytettävyyttä kollokaatiotutkimuksessa.

Tutkimusmenetelmä on yksikertainen ja ohjelmat vapaasti saatavilla

Rajoituksineenkin menetelmä, jossa perusmuotoistetusta aineistosta lasketaan esiintymisfrekvenssit, ja jossa konkordanssit tuotetaan pintamuotoisesta aineistosta perusmuodon perusteella ja tulostiedostot lajitellaan positiokohtaisesti, tarjoaa riittävän laajan materiaalin laadulliseen tarkasteluun. Mahdolliset homograafiset tai polyseemiset esiintymät havaitaan suhteellisen nopeasti tutkimustulosten joukossa ja on mahdollista tarkastella esiintymiä lause- tai virkerajojen yli. Samalla olen voinut vain pinnallisesti arvioida myös esiintymien semanttista motiivointia, kuten semanttista preferenssiä tai semanttista prosodiaa. Isommalla aineistolla tulosten siivoaminen käsin ei olisi mahdollista, joten tarvitaan merkityksen disambiguointia.

Jos tutkimus nyt aloitettaisiin alusta, se tehtäisiin eri tavoin aineiston käsittelyn, tulosten tuottamisen ja niiden arvioinnin osalta. Aineiston käsittelyssä hyödynnettäisiin dependenssijäsennettyä aineistoa manuaalisen työn sijaan ja tilastollisia menetelmiä käytettäisiin sekä tutkimuskohteen valinnassa että tulosten validoinnissa enemmän hyödyksi. Tarkastelen näitä seuraavaksi tarkemmin.

Tilastollisen analyysin hyödyntäminen

Tutkimusmenetelmä nojaa esiintymistiheyden ja laadullisen tarkastelun varaan. Tilastollisia menetelmiä on käytetty ainoastaan tutkimuskohteiden valinnan apuna ja satunnaisesti vahvistamaan havaintojen relevanssia, mutta havaintoja ei ole altistettu tilastolliselle laskennalle eikä tuloksista siten voi tehdä tilastollista päättelyä.

Valitsin verbit tutkimuskohteiksi tilastollisen merkitsevyyden perusteella, alkuvaiheessa WordSmithTools-ohjelman avulla ja myöhemmin NSP-ohjelman tuotta-

mien tunnuslukujen perusteella (ks. luku 5). Ammattikieli on kuitenkin niin tiivistä (ks. tyyppi/sananmuoto-tunnusluvut aluvussa 4.1), että aineistossa aivan liian monet yhteisesiintymät osoittautuvat tilastollisesti merkitseviksi, joten tuloslistoista syntyy helposti liian pitkiä manuaalisesti tutkittaviksi. Esimerkiksi NSP-ohjelmassa käytettävissä oleva Fisherin eksakti testi löysi varsinaisesta tutkimusaineistosta merkitseviä bigrammeja viiden saneen laajuisella laskentaikkunalla yli 60 000 riviä. Ratkaisuna voisi olla merkitsevyyden kynnyksarvon nostaminen, mutta se tuottaa lähinnä suljettujen leksikaalisten luokkien jäsenten yliedustusta tuloksissa, jolloin monet mielenkiintoiset havainnot jäävät pois tuloksista. Tässä tutkimuksessa asia on ratkaistu pääosin tilastollisen merkitsevyyden ohittamisella keskittyen puhtaisiin frekvensseihin ja niistä laskettuihin tunnuslukuihin (normalisoidut frekvenssit ja prosenttiosuudet). Tutkimuskohteiden esiintymistiheyksiä EU-aineistossa ja verrannaisaineistoissa on lisäksi tarkasteltu Khiin neliö -testin merkitsevyysarvojen avulla. Osa havaituista eroista todentuu merkitseväksi myös merkitsevyysarvojen perusteella ja voidaan yleisesti todeta, että tulokset vastaavat frekvenssien perusteella tehtyjä johtopäätöksiä.

Osa tutkimusaineiston havainnoista on yksittäisiä esiintymiä, jotka prosentuaalisesti tarkasteltuna tuottavat 100 prosentin osuuden kyseisen verbin ja tutkimuskohteenä olevan substantiivin yhteisesiintymistä, mutta joistakin on vain yksi havainto. Yksittäiset havainnot voivat vahvistaa muita samansuuntaisia havaintoja, tai ne voivat olla evidenssiä odotusten mukaisesta tai vastaisesta ilmiöstä kielessä, mutta ne eivät toimi todisteena yhteisesiintyjän kiinnittymisestä tiettyyn positioon.

(Dependenssi)jäsennyksen hyödyntäminen

Conexorin FinLite-ohjelmalla vuonna 2001 tehdyn käsittelyn tavoitteena oli saada aineisto perusmuoistetuksi saneiden laskentaa ja Unix-komennoilla käsittelyä varten. Dependenssijäsennystä ei silloin tehty, koska ajatuksena ei ollut hakea aineistosta yhteisesiintymiä syntaktisin kriteerein tai dependenssin perusteella. Tämä ratkaisu osoittautui harkitsemattomaksi, vaikkakin jäsennyks olisi myöhemminkin ollut mahdollista tehdä FDG-jäsentimellä esimerkiksi Kielipankissa olevaa jäsennintä käyttäen. Koska hakusanat ja yhteisesiintyjäehdokkaat oli jo valittu, olisi jäsennyks helpottanut käsityönä tehtävää analyysiä ja epäkelvojen hakutulosten poistaminen olisi jäänyt pois prosessista. Lisäksi halusin pitää tutkimuksen mahdollisimman aineistolähtöisenä, ja jos haun kriteerinä olisi käytetty yhteisesiintymien sanaluokkia, olisivat verbikantaiset substantiivit mahdollisesti jääneet pois tulosten joukosta. Jäsennetystä aineistosta olisi kuitenkin ollut hyötyä tutkimuksessa, ja jäsennyks kannattaisi jatkossa tehdä jo pelkästään tulosten tarkistamista helpottamaan. Varsinainen tutkimusaineisto ja ensimmäinen verrannaisaineisto jäsennettiin lopulta, jotta aineistosta saatiin selville tutkimusverbien esiintymistajuudet tilastollisten tunnus-

lukujen tuottamista varten.

Dependenssijäsennyksen tuottamia tuloksia olisi mahdollista arvioida tilastollisten tunnuslukujen avulla myös varsinaisessa tutkimusaineistossa ja ensimmäisessä verrannaisaineistossa. Tutkimusaineiston käsittelyssä on kuitenkin hyödynnetty ainoastaan lemmausta ja kaikki tutkimusikkunassa havaitut tutkimuskohteet on käsitelty ilman lähtötietoa dependensseistä. Tutkimusikkunaksi on valittu 11 saneen joukko, virkerajoja ei ole hyödynnetty eikä ole käytetty stop-listaa muun muassa välimerkkien ja numeroiden ohittamiseksi. Myöskään tilastollisia tunnuslukuja ei ole hyödynnetty tutkimuskohteiden valintavaihetta lukuunottamatta. Jäsennyksestä lähtevä tarkastelu sopii tutkimukseen, jossa tutkimuskohteiden syntaktiset roolit tiedetään etukäteen ja niiden avulla halutaan rajata muut havainnot tuloksista. Tämä ei ollut alkuperäisen tutkimusasetelman tavoitteena.

Toinen, kolmas ja neljäs verrannaisaineisto lisättiin tutkimusaineistojen joukkoon myöhemmin, koska tavoitteena oli saada tutkimusaineiston ja ensimmäisen verrannaisaineiston tuloksia peilaavaa materiaalia toisaalta uudemmassa käännytyssä ja alkukielisessä säädöskielessä ja toisaalta yleiskielessä. Nämä aineistot julkaistiin tutkijoiden käyttöön Kielipankissa myöhemmin, yleiskielinen aikakaus- ja sanomalehtiaineisto vuonna 2017 (neljäs verrannaisaineisto) ja alkuperäiset säädökset sisältävä aineisto (toinen verrannaisaineisto) kesällä 2019.

8.2 Ajatuksia tutkimuksen merkityksestä ja jatkotutkimusaiheista

Uusien potentiaalisten hakusanojen tutkimus ja ilmiön kuvauksen tarkentaminen

Pienessä mittakaavassa potentiaalisten hakusanojen tutkimusta olisi mahdollista ja hyödyllistä laajentaa myös tämän tutkimuksen aineistolla. Tilastollinen analyysi (NSP), jonka perusteella tutkittavat substantiivit ja potentiaaliset verbikollokaatit valittiin tutkimuskohteiksi, tarjosi myös muita mahdollisia tutkimuskohteita substantiivien **toimenpide**, **toimi**, **toiminta** ja **toiminto** lisäksi. Näitä ovat **hanke**, **menetelmä**, **menettely** ja **järjestely**, ja niiden tutkimus toisi lisävalaistusta tämän tutkimuksen tarjoamiin havaintoihin ja laajentaisi osaltaan ymmärrystä yhteisesiintymän olemuksesta ja merkityksestä säädöskielen rakentumisessa. Vakiolausekkeiden ja vakiolauseiden diakroninen tutkimus Kielipankin Semfinlex-kokoelman isoilla tekstiaineistoilla ja tutkimuksen ulottaminen muille oikeuskielen osa-alueille olisi käännöstyökalujen kehittämisen että lainsäätäjän ja lain tulkitsijan kannalta tarpeellista.

Syitä EU-aineiston ja SL-aineiston yhteisesiintymien eroavaisuuksien taustalla

Tutkimuksen taustatiedoiksi epätyypillisten kollokaatioiden syiden selvittämiseksi olisi tarvittu tietoa korpuksen koostamisen aikaan käytetyistä apuvälineistä, kääntäjän saamista ohjeista ja valmiista termistöistä. Lähtökieli ei ollut tiedossa. Ilman tietoa tilannekontekstista ei tuotettujen tekstien (käännösten, suomennosten tai suomenkielisten versioiden) perusteella ei tämän tutkimuksen puitteissa ollut mahdollista arvioida käännösprosessiin liittyviä tekijöitä ja syitä kääntäjien valintojen taustalla. Kysymystä on kuitenkin pyritty tarkastelemaan pohtimalla tutkittavan termin *toimenpide* mahdollisia käännösvastineita IATE-termipankissa, jonka perusteella havaittiin termin tavattoman laaja käyttöala. Termin **toimenpide** esiintymien määrä EU-aineistossa verrattuna verrannaisaineistoihin sai ainakin osittain selityksensä. Havainnon perusteella voidaan sanoa, että säädöskielessä jo käytössä oleville termeille todennäköisesti syntyy EU:n tarpeisiin uusia merkityksiä ja niille avautuu uudenlaisia käyttötapoja. Tätä ulottuvuutta olisi hyvä tutkia tarkemmin jatkossa.

Laadullisiin merkitykseen liittyviin ja toisaalta määrällisiin eroavaisuuksiin on hahmoteltavissa muutamia mahdollisia syitä. On osoitettu, että EU-kieli on oma kielivarianttinsa Koskinen (2000b), Westerlund (2006b), Piehl (2006), mikä selittäisi verrannaisaineistossa ennestään olemassa olevan termin esiintymisen täysin uudenlaisten yhteisesiintymien osana ja sitä kautta uudenlaisen merkityksen muodostumisen. Se selittäisi myös täysin uuden aineksen ilmaantumisen suomalaiseseen säädöskieleen, jolloin asetuksia tulkittaessa on syytä erottaa toisistaan vanhat merkityssisällöt ja uudet EU-tekstien myötä saapuneet merkitykset.

Kansainväliset yhteisöt, kuten Euroopan neuvosto tai YK tuottavat hybridikieltä Trosborg (1997) (ks. alaluku 3.5). Suomalaisessa säädöskielessä näkyy jälkiä myös yhteisesiintymien idiomaattisuudessa muista lähteistä saapuneista uudenlaisista elementeistä. Jotkin EU-tekstien kanssa samanlaiset yhteisesiintymärakenteet voidaan jäljittää esimerkiksi muiden kansainvälisten organisaatioiden tuottamiin teksteihin.

Toinen mahdollinen selitysmalli liittyy käännösprosessiin ja tarkoittaisi lähtökielen interferenssiä tulokielen yhteisesiintymärakenteissa joko yhteisesiintymän merkityksen muodostumisessa tai yhteisesiintyjien jakaumallisissa eroissa. Vaikka tutkimukseni osoittaa kiistattomia jakaumallisia eroja yhteisesiintyjien esiintymisessä tutkittujen yhteisesiintymien osana, ei ilman tietoa lähtökielestä ole mahdollista osoittaa interferenssin vaikutusta. Arvelin myös, että kaukana pääsanasta lähtökielessä sijaitseva kollokaatti voisi potentiaalisesti aiheuttaa epätyypillisen valinnan käännökseen Positiokohtainen analyysi ei kuitenkaan tue tätä ajatusta. Sen sijaan korostuu vakiolausekkeiden ja vakiolauseiden merkitys säädöstekstin käännöstyössä ja niiden systemaattiseen dokumentointiin olisi tutkimuksen perusteella tarvetta.

Kolmas selitysvaihtoehto liittyy kääntäjän tilannekontekstiin, toisin sanoen kaikkiin niihin reunaehtoihin, joiden varassa hän työtä tekee. Termipankit, käänösmuistit ja jo tehdyt käänökset sekä ohjeet ja säännöt tekevät kääntäjän puolesta valintoja, eikä EU-kontekstissa yksittäinen kääntäjä voi toimia poikkeavasti tai normien vastaisesti. Tulosten valossa näyttäisi siltä, että EU instituutioon tukisi adekvaattiin käänökseen tähtäävää käänösstrategiaa, joka painottaa lähtökielen ja -kulttuurin ensisijaisuutta.

Yhteisesiintymien elementtien paradigmaattinen vaihtelu, synonymia tai homonymia tuottavat haasteita yhteisesiintymän käytännön tutkimukselle. Tutkimukseni perusteella säädöskieli ammattikielenä sallii enemmän vaihtelua kuin mitä olisin odottanut löytäväni. Synonymian kartoittamiseen tarvitaan laajoja monikielisiä aineistoja ja semanttisen tiedon tietovarantoja.

Termipankkien ja sanastojen kehittäminen

Tutkimuksen valossa Pearsonin näkemys termien ei-systemaattisesta käytöstä (ks. alaluku 3.3) toteutuu suomalaisen lainsäädännön laajentuessa ja uusien käsitteiden määrän lisääntyessä. Kollokaatiotutkimuksen keinoin erikoisalan kielivarantoja kasvattamalla on mahdollista lisätä tietoisuutta alan käsitteistön todellisesta käytöstä suhteessa ohjeistukseen.

Kollokaatioiden tunnistaminen kääntäjän ja terminologin työvälineiksi olisi tuottava paremmin esille termipankkien ja sanastojen kehittämisessä. Erikoiskielten aineistojen koostaminen ja tutkimus toisi tähän hyviä mahdollisuuksia. Vakiolausekkeiden kattava dokumentointi olisi hyödyllistä erityisesti erikoisalojen kääntäjälle ja muillekin ammatissaan kieltä tuottaville, ja niitä olisi hyvä sisällyttää alan sanakirjoihin ja termipankkeihin. EU:n IATE-termipankki (www.iate.europa.eu) tarjoaa eri aloilta käänösvastineita ja autenttisia esimerkkejä termeistä kontekstissaan ja linkin tekstin kohtaan, jossa vastine esiintyy, sekä tarkat tiedot tekstistä. Tiedot sisältävät myös jonkinlaisen relevanssia osoittavan tähtiluokituksen, mutta tilastotietoja vastineen yleisyydestä tietopankista ei löydy. Hallinnon ja säädöstekstien tuottajat hyötyisivät ajantasaisista referenssiaineistoista.

Yhdyssanojen erikoislaadun huomioiminen ensimmäisen asteen kollokaationa olisi hyvä varmistaa termipankeissa ja kääntäjän käyttämissä sanastoissa ja sanakirjoissa. Erikois- ja ammattikielissä vakiintunut yhdyssana on termi, jonka merkitys rajaa sen kanssa käyttökelpoisten yhteisesiintyjien tai vakiintumien valikoimaa, eikä yhdyssanan edusosa aina määrittele tämän valikoiman kirjoa tai yhdyssanan kollokaatioprofilia. On myös mahdollista, että *toimenpide*-loppuisen yhdyssanan yhteisesiintyjä voi valikoitua sen määriteosan perusteella ja *toimenpide* voi osoittautua ylimääräiseksi yhdyssanan komponentiksi, jolla ei välttämättä ole mitään muuta

roolia merkityksen muodostumisessa kuin lisätä tekstin virallisuutta.

Kääntäjän tietoisuus kollokaatiosta ilmiönä ja sen eri olomuodoista

Kääntäjän on välttämätöntä tarkastella kieltä kontekstissaan, käyttää alun perin kohdekielellä kirjoitettuja verrannaismateriaaleja ja kiinnittää niissä erityistä huomiota tarkasteltavan termin tekstiyhteydessä yhteisesiintyviin elementteihin. Suomenkielisten lakitekstien aineistoja on jo saatavissa laajamittaiseen tutkimukseen ja niiden hyödyntäminen on perusteltua, mutta ne ovat Kielipankissa myös jokaisen kääntäjän käytettävissä. Samoin konekäännöstä editoivan kääntäjän olisi hyvä kiinnittää yhteisesiintymiin huomiota ja peilata niitä oman kielitajunsa lisäksi saman erikoisalan ja tekstityypin verrannaistesteihin.

Tässä tutkimuksessa on tarkasteltu termiä *toimenpide* ja sen tekstiyhteydessä esiintyviä verbejä *ryhtyä*, *soveltaa*, *suorittaa*, *säätää*, *tehdä* ja *toteuttaa*. Tulosten perusteella kääntäjän olisi hyvä varmistaa termien valinta tarkastelemalla verrannaisteksteissä tai termipankissa niiden yhteydessä käytettäviä verbejä. Verbien valinnalla vahvistetaan tulkintaa, jossa termi voi saada sellaisiakin lisämerkityksiä tai merkityksen ulottuvuuksia, joilla on jo vanhastaan suomalaisessa säädöskielessä vaikiintuneempi termi olemassa. Näin on käynyt termin *toimenpide* kanssa, joka on saanut **säädökseen** tai **säännökseen** liittyvän erikoismerkityksen.

Tutkimus tukee Trosborgin näkemystä (ks. alaluku 3.5) EU-kielestä hybridikielenä, joka noudattaa omia kollokaatioiden muodostumisen lainalaisuuksiaan. Hybridikieli ei ole hyvä tai huono asia, mutta on tärkeää tunnistaa EU-kielen olemus osana suomalaista säädöskieltä ja ottaa sen vaikutukset termien merkitykseen kontekstissaan huomioon säädöksiä tulkittaessa.

8.3 Kollokaatio kääntäjän työkaluksi

Kun tutkittava ilmiö ei ole ennen tutkimukseen ryhtymistä tarkkaan rajattu ja määritelty, on odotettavissa, että tutkimuksen kuluessa tapahtuu käsitteellistä tarkennusta ja tutkijan käsitys ilmiön luonteesta muovautuu havaintojen myötä. Näin kävi omassa tutkimuksessani. Alkuun pidin ortografista sanaa merkitysyksikkönä, jonka kombinatoriikkaa muiden samanlaisten kielen elementtien kanssa tulisin tutkimaan. Olin kiinnostunut muodon esiintymistä ja alkuun pohdin pitkään siirtymistä perusmuotojen tarkasteluun pintamuotojen sijaan. Työn edetessä ja havaintojen lisääntyessä fokus siirtyi muodosta merkityksen tarkasteluun. Siihen asti selkeinä pitämäni kielelliset elementit alkoivat menettää rajojaan ja sulautua erilaisiksi yhdistelmiksi, joilla näytti olevan merkitystä Firthin ja Sinclairin ajatuksia mukaillen yhdistelminä

ja ortografista sanaa laajempina kokoonpanoina.

Tämän tutkimuksen puitteissa pääsin tutkimaan kollokaation reaalistumia tekstissä kahdesta elementistä muodostuvan kollokaation tasolla. Kun päätös tutkimuskohteesta (substantiivi *toimenpide*) oli tehty, etsin aineistosta niiden kanssa usein esiintyviä verbejä. Näiden verbien keskinäisiä esiintymistajuuksien vertaaminen suomennoksissa alun perin suomeksi kirjoitetuista teksteistä koostuvista aineistoista tehtyihin laskelmiin. Ajatus ilmiön sisäkkäisyydestä ja syklistä luonteesta vahvistui useammasta kuin yhdestä elementistä koostuvien yhteisesiintymien myötä.

Konkreettisella tasolla ajattelun kehittyminen näkyy parhaiten ilmiölle oleellisten ominaispiirteiden hyväksymisenä työn lähtökohdaksi ja irtautumisena ortografian vaikutuspiiristä. Tutkittava ilmiö saattoi siis olla kahden tai useamman ortografisen sanan yhdistelmä, yhdyssana, tai se saattoi myös olla useamman ortografisen sanan muodostama lauseke. Käytännössä olio laajenee näin käsittämään kaiken materiaalin, kun ilmiö löytyy laajan kielellisen elementin sisältä ja tutkittava aines on kerrostunutta: se on sisäkkäistä ja limittäistä. Käsitteellisesti abstrakti ja monisyinen olio, kollokaatio, ei yhdessä tutkimuksessa tule kuvatuksi ja on tehtävä valintoja laadullisen analyysin asettamissa rajoissa. Tässä tutkimuksessa kohde onkin jäänyt paljon suppeammaksi kuin mihin käsitteen laajuus antaisi aiheita.

Kollokaation ominaispiirteet alkoivat työn kuluessa tarkentua näkymiksi samaan ilmiöön. Näkökulmat eivät ole toisiaan poissulkevia, mutta ne jakavat kollokaation käsitettä eri tavoin. Ollakseen käyttökelpoinen kääntäjälle, on käsite kuitenkin pysyttävä kuvaamaan tarkemmin. Hahmottelen tässä alaluvussa orastavan ajatuksen kollokaatiosta kääntäjän työkaluna, joka toimisi käytännön työskentelyssä varmistamassa kohdekielen odotuksenmukaisuutta, milloin se on kääntäjän tavoitteena. Hahmotelman pohjalla on ajatus jatkumosta, jossa yhteisesiintymä ja vakiintuma voivat ottaa useita sijainteja riippuen siitä, mikä näkökulma on aktiivinen. Yhteisesiintymä ja vakiintuma voivat siten edustua tekstissä samana merkkijonona.

Intensionaalisessa vaihtoehdossa lähdetään liikkeelle merkityksestä, distribuutiivisessa sanojen määrällisestä esiintymisestä ja esiintymien jakaumista. Kuvaan seuraavaksi näiden välisen jatkumon tutkimuksen perusteella ja näkökulmasta, joka on vastineeni Sinclairin abstraktiotason jatkumoon kollokaatiosta kolligaation kautta semanttiseen preferenssiin edelleen semanttiseen prosodiaan (Kübler ja Volanschi (2012)).

Ensin mainitulla intensionaalisella puolella kääntäjällä on tukiverkosto jo nykyisin: laadukkaat termipankit ja käännösmuistit tukevat yhteisesti sovittujen vastineiden valintaa, on käytettävissä selitysteoksia ja kielioppaita, kuten säädöskielen osalta oikeusministeriön Lainkirjoittajan opas⁶⁴ vuodelta 2013 ja Euroopan parlamentin,

⁶⁴https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/76493/lainkirjoittajan_opas_

neuvoston ja komission yhteisen käytännön opas Euroopan unionissa säädöstekstien laatimiseen osallistuvilla⁶⁵. Myös EU-kielenhuolto Kotimaisten kielten keskuksessa julkaisee EU-teksteihin liittyviä tutkimuksia ja ohjeita⁶⁶. Kääntäjän kannalta oleellista ovat toimintaa säätelevät normit ja suunnitelmallisuus tai sanavalintoja koskeva kääntäjän ulkoiset apuvälineet, termipankit ja käännösmuistit. Sanayhdistelmät jatkumon tässä päässä ovat **kiteymiä**, esimerkiksi täysin kiinteät vakiolausekkeet. Normien rikkominen kuvastaa tietämättömyyttä ja on siten tiedostamatonta. Käyttäjää tukevien apuvälineiden laadulla on oleellinen merkitys. Jatkumon tälle päälle tyypillistä on tietoisuus, **aie** tai **aikomus** käyttää sanavalintoja tietoisella tavalla.

Jatkumon toisessa päässä sijaitsee avoimen valinnan maailma, joka tämän tutkimuksen näkökulmasta on vähemmän vapaata kuin mitä on totuttu ajattelemaan, tosin jo säädöskieli materiaalina sisältää oletuksen vapaan valinnan vähäisyydestä. Avoin valinta ei ole myöskään kääntäjän työkalun hahmottamisen kannalta kovin kiinnostava. Kollokaation ilmenemismuoto, tapaan perustuva sanayhdistelmien rakentaminen on sen sijaan kääntäjälle haastava työtila, koska dokumentaatiota ja normistoa ei ole käytettävissä. Kääntäjä on oman kokemuksensa varassa valitessaan enemmän tai vähemmän tietoisesti sanayhdistelmien elementtejä.

Avoimesta valinnasta idiomaattisen valinnan suuntaan liikuttaessa tullaan **yhteisesiintymän** alueelle. Idiomaattinen yhteisesiintymä tuottaa kuitenkin lukijan tai kuulijan kannalta helpommin omaksuttavan tuloksen, joten asiaan on hyvä kiinnittää huomiota. Sanavalinnoille ei välttämättä löydy, joten olisi hyvä pohtia, onko esimerkiksi adjektiivin tai verbin valinta tekeillä olevan tekstityypin kannalta oleellinen kysymys vai ei. Jos se nähdään oleellisena kysymyksenä, tulee löytää keinot lisätä kääntäjän työkaluihin tietoa yhteisesiintymistä. Yksittäiselle yhteisesiintymälle ei aina löydy selitystä tai se on niin kaukana kielen kehityksen historiassa, ettei sitä enää osata sanallistaa. Tärkein este yhteisesiintyien sanallistamiselle kuitenkin on, että yksilö ei aina pysty tietoisesti tuottamaan tai arvioimaan niitä. Kääntäjän työn kannalta yhteisesiintyminen ilmiönä edellyttää valmiin tekstin tarkastelua ja huomion kiinnittämistä merkityksen rakentumiseen laajempien sanayhdistelmien kautta. Niissä yhdistelmän elementtien valinnalla on kokonaisuuden merkitystä tuottava tehtävä.

Kiteymä on oman tutkimukseni käsitteistön termein kiinteä kollokaatio. Kollokaation kompositionaalisuus ja läpinäkyvyys vähenevät kiteymää lähestyttäessä. Siirryttäessä kollokaatioon, jonka elementit sallivat eriasteista vaihtelua, puhun **vakiintumasta**. Vakiintuma edellyttää paitsi tietoa variaatiosta myös enemmän tietoisuutta sanayhdistelmien elementtien rajoituksista. Sen myötä suunnitelmallisuuden

low_20130904.pdf

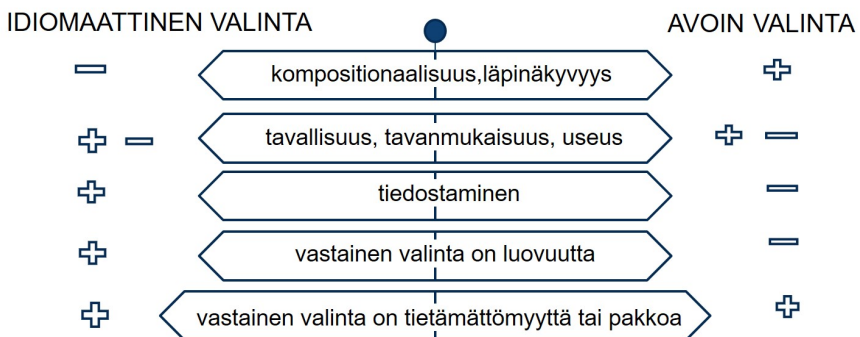
⁶⁵<https://eur-lex.europa.eu/content/techleg/FI-Saadostekstien-laadintaopas.pdf>

⁶⁶https://www.kotus.fi/ohjeet/saadoscielen_ohjeita

tarve kasvaa.

Merkityslähtöisen näkökulman eräs ulottuvuus on **myötäesiintymä**. Se sijoittuu avoimen valinnan, yhteisesiintymisen, vakiintumisen ja kiteytymisen kanssa samalle jatkumolle. Sillä tarkoitetaan syntaktisin kriteerein arvioitavaa sanasemantiikan asettamista rajoituksista johtuvaa kollokaatiota: esimerkiksi verbi **juoda** edellyttää kohteeltaan nestemäisyyttä. Myötäesiintymät liittyvät maailmantietoon ja laajempaan tilannekontekstiin ja ihmisen toiminnan ja fyysisen todellisuuden lainalaisuudet ovat tavallisesti kääntäjän näkökulmasta yleismaailmallisia, ja sanayhdistelmät ovat kompositionaalisia.

Vakiintuman ja yhteisesiintyjän alueilla on eniten kääntäjälle päänvaivaa aiheuttavia ilmiöitä. Mistä voi tietää, mikä elementti kyseisessä sanayhdistelmässä tietynä aikana ja tietyssä tekstilajissa on tavallinen tai konventionaalinen tai millaisia piilossa olevia konnotaatioita yhdistelmän valitseminen herättää? Miten oppia epäilemään omia sanavalintojaan? Yhteisesiintymän valinnassa määräävin elementti on tapa, jonka perusteella tehtyjä valintoja voidaan laskea tarkastelemalla toteutuneita valintoja. Vakiintumia ja kiteymiä voidaan samoin määritellä laskennallisin perustein. Vakiintumaan liittyy myös semanttisia piirteitä. Alla olevassa kuvassa 8.1 pyrin hahmottamaan yhteisesiintymän ja vakiintuman ominaisuuksien sijoittumista jatkumolle avoimesta valinnasta idiomaattisen valintaan.



Kuva 8.1: Yhteisesiintymän ja vakiintuman ominaisuuksien sijoittuminen jatkumolle idiomaattisen ja avoimen valinnan välillä.

8.4 Lopuksi

Yhteisesiintymän käsitteellinen rajausta ja tutkimusmenetelmät ovat kääntämisen tavoin tilannekontekstiinsa sidottua. Esimerkiksi osien lukumäärä tai osien välissä olevien elementtien sallittu määrä voi riippua kieliteknologisista mahdollisuuksista ja

merkityksen muodostumisen arviointi taustalla olevasta kielitieteellisestä teoriasta.

Kielen käyttäjälle tai ammattilaiselle hyödyllisenä käsitteenä yhteisesiintymä tulee voida tunnistaa ja kielelle tyypillisiä yhteisesiintymiä tulee voida kirjata muistiin ja hyödyntää referenssiaineistona apuvälineitä, termipankkeja tai muita sovelluksia tehtäessä. Kääntäjän tulee voida tarkistaa esimerkiksi sanakirjasta tai termipankista, edellyttääkö jokin käännösvastine muiden yhteisesiintymän osien valintaa ja mil-laista vaihtelun mahdollisuutta yhteisesiintymä tekstiin tai puheeseen rakentuessaan sallii.

Tämän tutkimuksen ilmiön rajaamiseen vaikuttavia tekijöitä ovat ammattikielen aineisto, tekstin jäsentämiseen liittyvät rajoitukset, käytettyjen kieliteknologisten menetelmien valinta ja kapasiteetti laadullisen analyysin läpiviemiseen. Nämä tilannekontekstiin liittyvät rajoitukset huomioiden tutkimus antaa viitteitä yhteisesiintymän ominaispiirteistä säädöskielessä.

Jos käännöstyön tavoitteena on, että käännös noudattaa suomalaisen säädöskielen konventioita ja integroituu jo olemassa olevaan säädöskokoelmaan, on kielen yhteisesiintymien hahmottaminen ja niiden noudattaminen oleellista. Jos taas tavoitteena on korostaa käännöksen alkuperää ja kääntämisen tilannekontekstia, esimerkiksi instituution erityispiirteitä, voi olla hyväkin, että käännös kuvastaa uudenlaisen säädöstyypin ominaispiirteitä poikkeamalla kohdekielen idiomaattisesta rakenteesta. Valinta on kuitenkin syytä tehdä tiedostaen. Myös suomalaisen säädöskielen systemaattisuutta on syytä tarkastella yhteisesiintymien tutkimuksen avulla tarkemmin. Säädöstekstien olisi erityisen suositeltavaa olla johdonmukaista ja odotuksenmukaista. Näin EU-suomi muuttuu kaikille suomalaisille ymmärrettäväksi.

Tutkimuksen valossa tutkimusaineistossa havaitut poikkeamat verrannaisaineiston yhteisesiintymissä kuvastavat Euroopan unionissa toimivan kääntäjän sidoksisuutta kielenulkoisiin olosuhteisiin, työvälineisiin ja kääntämisen reunaehtoihin, kuten odotuksiin tekstin lauserakenteen samankaltaisuudesta ja sovittuun terminologiaan sitoutumisesta. Terminologiasta vastaava yksikkö on haasteiden edessä, kun tarvitaan vastineita käsitteille, jotka eivät ole ilmiselvästi uusia mutta joita vanhastaan säädöskielessä käytössä oleva sanasto ei täsmälleen vastaa. Kun otetaan käyttöön olemassa olevaa terminologiaa, merkityskenttä laajenee ja syntyy uusia yhteisesiintymiä vanhojen rinnalle tai niitä ajan mittaan korvaamaan. Kääntämisen kautta säädöskielemme termien kanssa esiintyvien kielellisten elementtien jakaumat tulevat muuttumaan.

Kollokaatio ja sen eri ilmenemismuodot ominaispiirteineen ovat osoittautuneet haasteellisiksi, mutta sitäkin mielenkiintoisemmiksi tutkimuskohteiksi. Olen tutkimuksessani avannut näkökulmia kollokaation kontekstuaaliseen luonteeseen ja toivoakseni osoittanut sen kiinnostavuuden kielen mekanismien joukossa sekä merkityk-

sellisyyden odotuksenmukaisen ja idiomaattisen käännöksen tuottamisessa. Aloitin tutkimusmatkani J. R. Firthin parissa hyvän aikaa sitten ja hänen kanssaan (Palmer (1968): 182) se on hyvä päättää:

Collocations are actual words in habitual company. A word in a usual collocation stares you in the face just as it is.

Tämän tutkimukseni tavoitteena on ollut kurkistaa takaisin.

Aineistolähteet

EU

Westerlund, H. (1999). Kokoelma Euroopan yhteisöjen asetuksia. Eur-Lex.

SL

Kauppakaari (1999). Suomen säädöskokoelma. CD-ROM

semfinlex-lait

Helsingin yliopisto, FIN-CLARIN. Eduskunnan alkuperäissäädöksiä suomeksi vuosilta 1734-2018, korp-versio [tekstikorpus]. Kielipankki. Saatavilla <http://urn.fi/urn:nbn:fi:lb-2019042601>

LaDi

Kotimaisten kielten tutkimuskeskus (2016). Lakeja ja direktiivejä, Kielipankki-versio [tekstikorpus]. Kielipankki. Saatavilla <http://urn.fi/urn:nbn:fi:lb-20140730126>

lehdet-osa

Helsingin yliopisto (2017). 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä -korpus, versio 2 [tekstikorpus]. Kielipankki. Saatavilla <http://urn.fi/urn:nbn:fi:lb-2017091901>

Liitemateriaalia

Tekijältä on saatavissa seuraavat taustamateriaalit:

Käsitteistö

Lista varsinaisen tutkimusaineiston sisältämistä asetuksista

Lista yhdyssanoista EU- ja SL-aineistoissa ja niiden frekvenssit

Konkordanssit tutkimuskohteista EU- ja SL-aineistoissa

Aineistoista semfinlex-lait, LaDi ja lehdet-osa haut voi tuottaa Kielipankin työkaluilla

Lähteet

- Abney, S. (1996). *Statistical Methods and Linguistics*. Teoksessa Klavans, J. & Resnik, P. (toim.) *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. The MIT Press, Cambridge, MA.
- Alarotu, A. (2001). *Termien vakiintuminen - EU-termejä lehtikielessä*. Pro Gradu Thesis. Tampereen yliopisto, Kieli- ja käännöstieteiden laitos.
- Aro, J. (2012). *Aivoinfarktipotilaiden sanahaun prosessi kuntoutumisen akuuttivaiheessa*. Väitöskirja, Tampereen yliopisto. DOI: urn.fi/urn:nbn:fi:uta-1-22316
- Arppe, A. (2008). *Univariate, bivariate, and multivariate methods in corpus-based lexicography: a study of synonymy*. PhD thesis, Publications of the Department of General Linguistics, University of Helsinki, No. 44.
- Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge University Press, Cambridge.
- Baker, M. (1993). *Corpus Linguistics and Translation Studies. Implications and Applications*. Teoksessa Baker, M., Gill, F., Sinclair, J. & Tognini-Bonelli, E. (toim.) *Text and Technology. In Honour of John Sinclair*. John Benjamins, Amsterdam/Philadelphia.
- Baker, M. (1996). *Corpus-based translation studies: The challenges that lie ahead*. Teoksessa Somers, H. (toim.) *Terminology, LSP and Translation. Studies in language engineering in honour of Juan C. Sager*, s.175–186. John Benjamins, Amsterdam/Philadelphia.
- Baker, M. (2011). *In other words: a coursebook on translation*. Taylor & Francis, Oxford, 2nd edition.
- Baldwin, T. & Kim, S. N. (2010). *Multiword Expressions*. Teoksessa Indurkha, N. & Damerau, F. (toim.) *Handbook of Natural Language Processing*, s. 267–292. CRC Press, Boca Baton, 2nd edition.
- Banerjee, S. & Pedersen, T. (2003). *The Design Implementation and Use of the Ngram Statistics Package*. Teoksessa *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, s. 370–381.
- Bartsch, S. (2009). *Patterns of verb colligation and collocation in scientific discourse*. In Mahlberg, M., González-Díaz, V. & Smith, C. (toim.) *Patterns of verb colligation and collocation in scientific discourse. CL2009*. Liverpool. <http://ucrel.lancs.ac.uk/publications/cl2009/>

- Bernardini, S. (2015). Translation. *The Cambridge Handbook of English Corpus Linguistics*, s. 515–536. Cambridge University Press, Cambridge.
- Biber, D. (2015). *The Cambridge handbook of English corpus linguistics*. Cambridge University Press, New York.
- Bowker, L. & Pearson, J. (2002). *Working with Specialized Language*. Routledge, London.
- Chapman, S. & Routledge, C. (2005). *Key thinkers in linguistics and the philosophy of language*. Oxford University Press, New York.
- Charniak, E. (1997). *Statistical Language Learning*. MIT Press, Massachusetts.
- Chesterman, A. (1997). *Memes of Translation. The spread of ideas in translation theory*. John Benjamins Publishing, Amsterdam and Philadelphia.
- Chesterman, A. ja Wagner, E. (2002). *Can Theory Help Translators - A dialogue between the ivory tower and the wordface*. St. Jerome Publishing, Manchester and Northampton.
- Chesterman, A. (2004). Hypotheses about translation universals. Teoksessa Hansen, G., Malmkjær, K. & Gile, D. (toim.) *Claims Changes and Challenges in Translation Studies*. Selected Contributions from the EST congress Copenhagen 2011, s. 1–14. John Benjamins Publishing, Amsterdam and Philadelphia.
- Clear, J. (1993). From Firth Principles. Computational Tools of the Study of Collocation. *Text and Technology*. In *Honour of John Sinclair*. John Benjamins, Amsterdam/Philadelphia.
- Crystal, D. (2003) *A dictionary of linguistics and phonetics*. Blackwell, Oxford
- Eskola, S. (2002). *Syntetisoivat rakenteet ja käännessuomi: suomennetun kaunokirjallisuuden erityispiirteiden tarkastelua korpusmenetelmällä*. Väitöskirja, Joensuun yliopisto
- Evert, S. (2005). *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. PhD thesis, Institut für maschinelle Sprachverarbeitung, Universität Stuttgart
- Evert, S. (2008). Corpora and collocations. Teoksessa Lüdeling, A. & Kytö, M. (toim.) *Corpus Linguistics. An International Handbook.*, vol. 2, s. 1212–1248. Mouton de Gruyter, Berlin.
- Evert, S. & Krenn, B. (2003). Computational approaches to collocations. Introductory course at the European Summer School on Logic, Language, and Information (ESSLLI 2003). <http://www.collocations.de/EK/>

- Fellbaum, C. (2011). Idioms and collocations. Teoksessa Maienborn, C. & von Heusinger, K. & Portner, P. (toim.) *An International Handbook of Natural Language Meaning*, chapter 20, vol. 1, s. 441–456. John Benjamins.
- Firth, J. R. (1957). *Papers in linguistics 1934-1951*. Oxford University Press, London.
- Firth, J. R. ja Palmer, F. R. (1968). *Selected papers of J. R. Firth 1952-59*. Longmans, London.
- Genzler, E. (2001). *Contemporary Translation Theories*. Multilingual Matters, Clevedon.
- Gledhill, C. (1996). The phraseology of rhetoric, collocations and discourse in cancer research abstracts. Teoksessa Barron, C. & Bruce, N. (toim.) *Knowledge and Discourse: Proceedings of the International Multidisciplinary Conference, 1996*. Hong-Kong, China. <http://ec.hku.hk/kd96proc/authors/papers/Gledhill.htm>
- Gledhill, C. (2000). Collocations in Science Writing. Teoksessa *Language in Performance Series*. vol. 22 Gunter Narr Verlag, Tübingen.
- Granger, S. ja Paquot, M. (2008). Disentangling the phraseological web. Teoksessa Granger, S. ja Meunier, F. (toim.) *Phraseology: An Interdisciplinary Perspective*, s. 27–49. John Benjamins Publishing Co, Amsterdam/Philadelphia.
- Gray, B. ja Biber, D. (2015). Phraseology. Teoksessa Biber, D. & Reppen, R. (toim.) *The Cambridge Handbook of English Corpus Linguistics*, s. 125–145. Cambridge University Press, Cambridge.
- Gries, S. T. (2008). Phraseology and linguistic theory. A brief survey. Teoksessa Granger, S. ja Meunier, F. (toim.) *Phraseology: An Interdisciplinary Perspective*, s. 3–25. John Benjamins Publishing Co, Amsterdam/Philadelphia.
- Gries, S. T. (2013). 50-something years of work on collocations: what is or should be next... *International Journal of Corpus Linguistics*, 18(1):137–165. http://www.linguistics.ucsb.edu/faculty/stgries/research/2013_STG_DeltaP\&H_IJCL.pdf
- Gries, S. T. (2015). Quantitative designs and statistical techniques. Teoksessa Biber, D. & Reppen, R. (toim.) *The Cambridge Handbook of English Corpus Linguistics*, s. 50–72. Cambridge University Press, Cambridge.
- Hakulinen, A. ja Ojanen, J. (1970). *Kielitieteen ja fonetiikan terminologiaa*. Suomalaisen Kirjallisuuden Seura, Helsinki.
- Hakulinen, A., Vilkuna, M., Korhonen, R., Koivisto, V., Heinonen, T. R., ja Alho, I. (2004). *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seura, Helsinki.

- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold, London.
- Halliday, M. A. K. (1992). Language as a system and language as instance: the corpus as a theoretical construct. Teoksessa Svartvik J. (toim.) *Directions in Corpus Linguistics*, s. 61–77. Mouton, Berlin.
- Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. Arnold, London.
- Halliday, M. A. K. ja Hasan, R. (1985). *Language, context, and text: aspects of language in a social-semiotic perspective*. Oxford University Press, Oxford.
- Hanks, P. (2008). The Lexicographical Legacy of John Sinclair. *International Journal of Lexicography*, 21(3):219–229.
- Hanks, P. (2010). Lexicography, Terminology, and Phraseology. In Proceedings of Euralex 2010, *Leeuwarden*.
- Hanks, P. (2012). How people use words to make meanings: Semantic types meet valencies. *Input, Process and Product: Developments in Teaching and Language Corpora*. Masaryk University Press.
- Hausmann, F.-J. (2004). *Was sind eigentlich Kollokationen? Wortverbindungen - mehr oder weniger fest*, s. 309–334. Institut für Deutsche Sprache, Jahrbuch 2004.
- Heid, U. (1999). Extracting Terminologically Relevant Collocations from German Technical Texts. In Sandrini, P., (toim.) *Proceedings Fifth International Congress on Terminology and Knowledge Engineering, August 1999 TKE99 Terminology and Knowledge Engineering*, s. 23–27, Vienna. TermNet.
- Heid, U. (2001). Collocations in Sublanguage Texts: Extraction from Corpora. Teoksessa Wright, S. E. and Budin, G. (toim.) *Handbook of Terminology Management*, s. 788–808.
- Helin, I. (1998). Kollokationen als Bestandteil einer Fachsprache. Teoksessa Haukioja, T., (toim.), s. 177–184 *Proceedings of the XVIth Scandinavian Conference of Linguistics*. University of Turku.
- Herbst, T. (2012). Choosing sandy beachies - collocation, probabemes and the idiom principle. Teoksessa Herbst, T. & Faulhaber, S. & Uhrig, P. *The Phraseological View of Language: A Tribute to John Sinclair*, pages 39–57. Mouton De Gruyter, Berlin.
- Herbst, T. & Götz-Votteler, K. (2007). Introduction: the Mystery of Collocation. *Zeitschrift für Anglistik und Amerikanistik*, (55.3):211–215.

- Hermans, T. (1999). *Translation in Systems: Descriptive and System-Oriented Approaches Explained (Translation Theories Explored)*. St. Jerome Publishing, Manchester.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge, London.
- Honeybone, P. (2005). Firth J. R. (John Rupert). Teoksessa Routledge, s. & Routledge, P. (toim.) *Key Thinkers in Linguistics and the Philosophy of Language*, s. 80–86. Edinburg University Press, Edinburg.
- Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, (12:2):249–268.
- Inkpen, D. Z. ja Hirst, G. (2002). Acquiring Collocations for Lexical Choice between Near-Synonyms. Teoksessa *Unsupervised Lexical Acquisition: Proceedings of the Workshop of the ACL Special Interest Group on the Lexicon (SIGLEX)*, s. 67–76.
- Jantunen, J. H. (2001). ”Tärkeä” seikka ja ”keskeinen” kysymys: Mitä korpuslingvistinen analyysi paljastaa lähisynonyymeista *Virittäjä*, 105(2):170. <http://urn.fi/URN:NBN:fi:ELE-865950>
- Jantunen, J. H. (2004). *Synonymia ja käännessuomi: Korpusnäkökulma samamerki-tyksisyyden kontekstuaalisuuteen ja käännesskielen leksikaalisiin erityispiirteisiin*. PhD thesis, Joensuun yliopisto.
- Jantunen, J. H. (2009). Minulla on aivan paljon rahaa. Fraseologiset yksiköt suomen kielen opetuksessa. *Virittäjä*, 113 (3):356–381. Kotikielen seura, Helsinki.
- Jantunen, J. H. (2016). Corpora, phraseology and dictionaries: How does corpus research intersect language teaching and learning? Teoksessa Vilas. B. S. (toim.) *Collocations cross-linguistically: Corpora, dictionaries and language teaching*, s. 97–119. Uusfilologinen Yhdistys.
- Jantunen, J. H. ja Brunni, S. (2012). Morfologinen priming ja fraseologia vieraan kielen oppimisessa: Korpustutkimus oppijansuomesta. *Lahivordlusi Lahivertailuja*, 22:71–100.
- Jäppinen, U. (1999). *Kielen ja ajatusten kliseet. Värien nimitysten käyttö eräissä suomenkielisissä idealistissävyisissä teksteissä*, Väitöskirja, Åbo Akademi.
- Järvelin K. (1995) *Tekstitiedonhaku tietokannoista: Johdatus periaatteisiin ja menetelmiin*. Suomen atk-kustannus, Espoo.
- Järvelin K. ja Sormunen, E. (2010). Tiedon tallennus ja haku. Teoksessa Hentonen, P. & Serola, S. (toim.) *Ote informaatiosta. Johdatus informaatiotutkimukseen ja interaktiiviseen mediaan*. BTJ, Helsinki.

- Jääskeläinen, R. (1999). *Tapping the Process: An Explorative Study of the Cognitive and Affective Factors Involved in Translating*. PhD thesis, University of Joensuu.
- Karlsson, F. (2008). *Yleinen kielitiede*. Gaudeamus Helsinki University Press, Helsinki, Uud. laitoksen 3. p.
- Karvonen, P. (1997). *Suomi eurooppalaisessa kieliyhteisössä*. Opetusministeriö, Helsinki.
- Kemppanen, H. (2008). *Avainsanoja ja ideologiaa: Käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi*. Väitöskirja, Joensuun yliopisto.
- Kenny, D. (2001). *Lexis and creativity in translation: A corpus-based study*. St. Jerome Publishing, Manchester.
- Ketola, J. (2002). *Työoikeudellisten lakitekstien ideationaalisia, tekstuaalisia sekä intersubjektiveja merkityksiä*. Lisensiaatintyö, Vaasan yliopisto, Humanistinen tiedekunta, Nykysuomen ja kääntämisen laitos.
- Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory*, 1–2:263–275.
- Kinnunen, T. (2006). *Käännös oikeudenkäyntiaineistona: Lakitekstien tulkinta oikeuskäsitteiden kääntämisen näkökulmasta*. Väitöskirja, Tampereen yliopisto.
- Korhonen, J. A. (2018). *Fraseologia: kiinteiden sanayhtymien tutkimus*. Finn Lectura, Helsinki.
- Koskinen, K. (2000a). *Beyond Ambivalence. Postmodernity and the Ethics of Translation*. PhD thesis, University of Tampere.
- Koskinen, K. (2000b). Institutional Illusions. Translating in the EU Commission: A Finnish Perspective. *The Translator. Studies in Intercultural Communication*, 6(1):49–65.
- Krenn, B. ja Evert, S. (2001). Can we do better than frequency? A case study on extracting PP-verb collocations. In *Proceedings of the ACL Workshop on Collocation: Computational Extraction, Analysis and Exploitation*, s. 39–46. Association for Computational Linguistics.
- Kussmaul, P. & Tirkkonen-Condit, S. (1995). Think-Aloud Protocol Analysis in Translation Studies. *TTR: traduction, terminologie, rédaction*, 8(1):177–199.
- Kübler, N. & Volanschi, A. (2012). Semantic prosody and specialised translation, or how a lexicogrammatical theory of language can help with specialised translation. Teoksessa Boulton, A., Carter-Thomas S., & Rowley-Jolivet, E. (toim.) *Corpus-informed Research and Learning in ESP: Issues and Applications*, s.105–135. John Benjamins Publishing, Philadelphia/Amsterdam.

- Laviosa, S. (1996). *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study for Translation*. PhD thesis, Manchester, UMIST.
- Léon, J. (2007). From Linguistic Events and Restricted Languages to Registers. Firthian legacy and Corpus Linguistics. *Henry Sweet Society for the History of Linguistic Ideas Bulletin*, 49:5–25.
- Lehr, A. (1993). *Kollokationsanalysen - Von der Kollokationstheorie des Kontextualismus zu einem computergestützten Verfahren*. Walter de Gruyter, Berlin.
- Lewis, M. (2000). Language in the lexical approach. Teoksessa Lewis, M. (toim.) *Teaching Collocation: Further Developments In The Lexical Approach*, s. 126–154. Language Teaching Publications, Hove.
- Liebscher, R. ja Groppe, D. (2003). Rethinking context availability for concrete and abstract words: a corpus study. In *Proceedings of Corpus Linguistics 2003*.
- Luukka, M.-R. (2002). M.A.K. Halliday ja systeemis-funktionaalinen kielitiede. Teoksessa Dufva, H. & Lähteenmäki, M. (toim.) *Soveltavan kielentutkimuksen teoriaa ja käytäntöä*, 4:89–123. Soveltavan kielentutkimuksen keskus, Jyväskylä.
- Léon, J. (2007). From Linguistic Events and Restricted Languages to Registers. Firthian legacy and Corpus Linguistics. *Henry Sweet Society for the History of Linguistic Ideas Bulletin*, 49:5–25.
- Manning, C. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Massachusetts.
- Mattila, H. E. S. (2002). *Vertaileva oikeuslingvistiikka*. Lakimiesliiton kustannus Kauppakaari, Helsinki.
- Mattila, H. E. S. (2010). *Oikeuskieli ja säädöstieto. Suomenkielinen lakikirja 250 vuotta Rättspråk och författningsinformation. Den finskspråkiga lagboken 250 år*. Suomalainen Lakimiesyhdistys, Helsinki.
- Mauranen, A. (2000). Strange Strings in Translated Language. Teoksessa Olohan, M. *Intercultural Faultlines. Research Models in Translation Studies I: Textual and Cognitive Aspects*, s. 119–141. St. Jerome Publishing, Manchester.
- McKeown, K. R. ja Radeev, D. R. (1999). Collocations. Teoksessa Dale, R. & Moisl, H. & Harold Somers, H. (toim.) *A Handbook of Natural Language Processing*. Marcel Dekker.
- Mel'čuk, I. (1998). Collocations and Lexical Functions. Teoksessa Cowie, A. P. (toim.) *Phraseology. Theory, Analysis and Applications*, pages 23–53. Clarendon Press, Oxford.

- Mäkisalo, J. (2000). *Grammar and experimental evidence in Finnish compounds*, Väitöskirja, Joensuun yliopisto.
- Nordman, L. (2009). Lagöversättning som process och produkt. *Revideringar och institutioner vid lagöversättning till svenska I Finland*. Väitöskirja. Nordica Helsingiensis 16.
- Oakes, M. (1998). *Statistics for corpus linguistics*. Edinburgh University Press, Edinburgh.
- Palmer, F. R. (1968). *Selected Papers of J. R. Firth 1952-59*. Indiana University Press.
- Pandit, P. B. (1970). Review of F.R. Palmer (ed.): Selected papers of J.R. Firth 1952-1959. *Journal of linguistics*, (6):280–284.
- Partington, A. (1993). Corpus Evidence of Language Change. The Case of the Intensifier. Teoksessa Baker, M. & Francis, G. & Sinclair, J. M. & Tognini-Bonelli, E. (toim.) *Text and technology: In honour of John Sinclair*, s. 176–192. J. Benjamins Pub. Co., Philadelphia.
- Partington, A. (1998). *Patterns and Meanings. Using corpora for English language research and teaching*. John Benjamins, Amsterdam.
- Partington, A. (2004). ”Utterly content in each other’s company”: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, (9/1):131–156.
- Pasanen, P. (2009). *Merenkulun turvallisuuden koetinkiviä: Terminologisen tiedon poiminta teksteistä*. Väitöskirja, Helsingin yliopisto.
- Pearson, J. (1998). *Terms in context*. John Benjamins, Amsterdam.
- Pedersen, T. (2011). Identifying Collocations to Measure Compositionality: Shared Task System Description. In *Workshop on Distributional Semantics and Compositionality DiSCo’201, Portland, Oregon, 24 June 2011*.
- Piehl, A. (2003). ”Selkeyttä vain” - EU-kieleen. *Kielikello*, (4). www.kielikello.fi/index.php?mid=2&pid=11&aid=1461
- Piehl, A. (2006). The Influence of EU legislation on Finnish legal language. Teoksessa Gotti, M. & Giannoni, D. S. *New trends in specialized discourse analysis: Linguistic insights*, 44:183–194.
- Piehl, A. (2013). Fine tuning style and precision: Adapting directive citations to Finnish statutes. *Linguistica Antverpiensia, New Series – Themes in Translation Studies (LANS – TTS)*, Research models and methods in legal translation (12):161–181. <https://lans-tts.uantwerpen.be/index.php/LANS-TTS/issue/view/5>

- Probirskaja, S. (2009). *Rajankäyntiä. Suomen ja Venäjän kahdenväliset valtiosopi- mukset käännöstieteellisen avainsana-analyysin valossa*. Väitöskirja, Tampereen yliopisto.
- Rayson, P. (2002). *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. PhD thesis, Lancaster University.
- Rayson, P. (2015). Tools and methods for corpus compilation and analysis. Teoksessa Biber, D. and Reppen, R. (toim.) *The Cambridge Handbook of English Corpus Linguistics*, s.32–49. Cambridge University Press, Cambridge.
- Rundell, M. ja Kilgarriff, A. (2011). Automating the creation of dictionaries. where will it all end? In Granger, A. & Meunier, F. (toim.) *A Taste for Corpora: In Honour of Sylviane Grange.*, v. 45 in *Studies in Corpus Linguistics*, s. 257–281. John Benjamins Publishing Company.
- Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, R. (2002). Multiword expressions: a pain in the neck for NLP. *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, s.1–15.
- Saldanha, G. (2009). Principles of corpus linguistics and their application to translation studies research. *Revista Tradumática*, (7). 01/art.
- Sardinha, T. B. (2017). Lexical priming and register variation. *Studies in Corpus Linguistics*, s. 189–229. John Benjamins Publishing Company, Cambridge.
- Schäffner, C. (1997). *Translation and Norms*. Multilingual Matters, Clevedon.
- Schönfeld, D. (2011). Introduction. On evidence and the convergence of evidence in linguistic research. Teoksessa Schönfeld, D. (toim.) *Converging evidence: Methodological and theoretical issues for linguistic research*, (33):1–31. John Benjamins, Amsterdam & Philadelphia.
- Seretan, V. (2011). *Syntax-Based Collocation Extraction*, Volume 44. Springer
- Shore, S. (2004). J.R. Firth, ojenne luentosarjassa ”Kielitieteen lukemattomat klassikot” 28.4.2004.
- Shore, S. (2012). Kieli, kielenkäyttö ja kielenkäytön lajit systeemis-funktionaalisessa teoriassa, s. 131–157. Teoksessa Heikkinen, V. & Voutilainen, E. & Lauerma, P. & Tiililä, U. & Lounela, M. (toim.) *Genreanalyysi – tekstilajitutkimuksen käytäntöä*. Kotimaisten kielten keskus.
- Sinclair, J. (1988). Collocation: a progress report. Teoksessa Burton, T. L. and Burton, J. (toim.) *Lexicographical and Linguistic Studies, Essays in Honour of G. W. Turner*, s. 319–331. D.S. Brewer, Cambridge.

- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press, Oxford.
- Sinclair, J. (1996). The Search for Units of Meaning. *Textus*, (IX):75–106.
- Sinclair, J. ja Carter, R. (2004). *Trust the text. Language, corpus and discourse*. Routledge, London.
- Sinclair, J., Jones, M., Daley, S., & Krishnamurthy, R. (2004). *English collocation studies: The OSTI report*. Continuum, London.
- Stewart, D. (2009). Safeguarding the lexicogrammatical environment: Translating semantic prosody. Teoksessa Beeby, A., Rodriguez-Ines, P. & Sanchez-Gijon, P. (toim.) *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*, s. 29–46. John Benjamins, Amsterdam.
- Stubbs, M. (1995). Corpus Evidence for Norms of Lexical Collocation. Teoksessa Cook, G. & Seidlhofer B. (toim.) *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson*, s. 245–256. Oxford University Press, Oxford.
- Stubbs, M. (1996). *Text and corpus analysis: computer-assisted studies of language and culture*. Blackwell, Oxford.
- Stubbs, M. (1996). *Words and Phrases: Corpus Studies of studies of Lexical Semantics*. Blackwell, Oxford.
- Tapanainen, P. ja Järvinen, T. (1997). A non-projective dependency parser. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, s. 64–71. Washington D.C.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Number 6. John Benjamins Publishing Company, Amsterdam.
- Toury, G. (1995). *Descriptive Translation Studies - and beyond*. John Benjamins Publishing Company, Amsterdam/Philadelphia, 1st edition.
- Toury, G. (1999). A Handful of Paragraphs on "Translation norms" and "Norms". Teoksessa Schäffner, C. (toim.) *Translation and norms*, s. 9–31. Multilingual Matters, Clevedon.
- Toury, G. (2000). The Nature and Role of Norms in Translation. Venuti, L. (toim.) *The Translation Studies Reader*, s. 198–211. Routledge, London.
- Toury, G. (2004). Probabilistic explanations in translation studies. Welcome as they are, would they qualify as universals? Kujamäki, P. ja Mauranen, A. (toim.) *Translation Universals: Do They Exist?* s. 15–82. John Benjamins Publishing Company, Amsterdam/Philadelphia.

- Toury, G. (2012). *Descriptive Translation Studies - and beyond*. John Benjamins Publishing Company, Amsterdam/Philadelphia, revised edition.
- Trask, R. L. (1997). *A student's dictionary of language and linguistics*. Arnold, London.
- Trosborg, A. (1997). Translating Hybrid Political Texts. Trosborg, A. (toim.) *Text Typology and Translation*, s. 145–158. John Benjamins, Amsterdam and Philadelphia.
- Uosukainen, S. (2006). *Adjektiivihin liittyvät kollokaatiot kuntien internet-sivujen esittelyteksteissä. Korpuslingvististen menetelmien mahdollisuudet ja rajat kvalitatiivisessa tutkimuksessa*. Pro Gradu, Helsingin yliopisto.
- Vanhatalo, U. (2003). Kyselytestit vs. korpuslingvistiikka lähisynonyymien semanttisten sisältöjen arvioinnissa – mitä vielä keskeisestä ja tärkeästä? *Virittäjä*, (3): 351-369.
- Vilkaitė, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(10):1632–1642.
- Vilkaitė, L. & Schmitt, N. (2017). Reading Collocations in an L2: Do Collocation Processing Benefits Extend to Non-Adjacent Collocations? *Applied Linguistics*, 40(2):329–354.
- Westerlund, H. (2000). Leksikaalisista kollokaatioista EU-asetusten käännöksissä ja Suomen lainsäädännön kielessä. In *Erikoiskielet ja käännösteoria. VAKKI-symposiumi XX. Vaasa 11.-13.2.2000*, s.400–406.
- Westerlund, H. (2006a). Euroopan yhteisöjen asetukset kääntäjän haasteena. *Terminfo*, (4).
- Westerlund, H. (2006b). Säädoskielen kääntäjä ja yhdyssanan houkutus. Teoksessa Piehl, A. and Vihonen, I. (toim.) *Vuosikymmen EU-suomea*, s. 75–88. Kotimaisten kielten tutkimuskeskus, Helsinki, 142 edition.
- Westerlund, H. (2009). Towards defining collocation as a tool for the translator. Helin, I. (toim.) *Linguistik und Übersetzung in Kouvola: Beiträge zu Sprache und Sprachen 7: Vorträge der 17. Jahrestagung der Gesellschaft für Sprache und Sprachen (Jesus e.V.)*, (VI): 57–68. Universität Helsinki, Institut für Translationswissenschaft, Helsinki.
- Wright, S. E. (1997). Term Selection: The Initial Phase of Terminology Management. Teoksessa Wright, S. E. & Budin, G. (toim.) *Handbook of Terminology Management*: (1): 13–143 Basic Aspects of Terminology Management. John Benjamins Publishing Company, Amsterdam.

- Xiao, R. (2015). Collocation. Teoksessa Biber, D. & Reppen, R. (toim.) *The Cambridge Handbook of English Corpus Linguistics*, s. 106–124. Cambridge University Press, Cambridge.
- Yli-Jokipii, H. (2004). Tekniikan kieli tutkimuksen ja kääntämisen näkökulmasta. Teoksessa Tommola, J. (toim.) *Kieli, teksti ja kääntäminen. Language, Text and translation*, s. 81–94. Turun yliopisto, Turku.
- Yuret, D. (1998). *Discovery of Linguistic Relations Using Lexical Attraction*. PhD thesis, MIT Artificial Intelligence Laboratory.